

16	245.5	3.8	1487	16	Q8YK40	Q8yk40 anabaena sp
----	-------	-----	------	----	--------	--------------------

121 REDQROT RAYYALVND EANH LKRLNTINSNRIGNRNNNSKFVIGVDNPAHVLRFTDDG 180

```
Db      121 REDQRTYALVNDENVHLK----- 144
```

```

QY 181 TKFNFTNOTGGEIVNDFILDAPIILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSG 240
Db 145 ----- 144
QY 241 TNADGMPDGGNGQITNTDPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHKIKT 300
Db 145 ----- 144
QY 301 SFOLDERINTNSRIGNRRNNNSKFEYIGVDNPAHYIRFTDDGTFKFNFTNOTGGEIVNDFI 360
Db 145 ----- RINTNSRIGNRRNNNSKFEYIGVDNPAHYIRFTDDGTFKFNFTNOTGGEIVNDFI 198
QY 361 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNNADGMPDGGNGQITNT 420
Db 199 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNNADGMPDGGNGQITNT 258
QY 421 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHKIKTSFOLDERFVYPWTGSEE 480
Db 259 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHKIKTSFOLDERFVYPWTGSEE 318
QY 481 NKNTIRLATGSLPSENERWILDPGTQVTLKEDSVNVFSRLVNSLSFTGSIYIF 540
Db 319 NKNTIRLATGSLPSENERWILDPGTQVTLKEDSVNVFSRLVNSLSFTGSIYIF 378
QY 541 GTSELPSLWYSFPTRLSDLTALNOVKTTDIEASTDNGTTTGTADTSSGSGTACT 600
Db 379 GTSELPSLWYSFPTRLSDLTALNOVKTTDIEASTDNGTTTGTADTSSGSGTACT 438
QY 601 GMTNTSOTVSNPTLNTYRSFGIDSKPTSAKIDETWADPNVTEATVAYEYRGLGIONEL 660
Db 439 GMTNTSOTVSNPTLNTYRSFGIDSKPTSAKIDETWADPNVTEATVAYEYRGLGIONEL 498
QY 661 PITNAGNFTIRNTIGVGEFTSGSRVYLKASVNGDORPTGNFPLVYFGYLGYOQTRTGT 720
Db 499 PITNAGNFTIRNTIGVGEFTSGSRVYLKASVNGDORPTGNFPLVYFGYLGYOQTRTGT 558
QY 721 FWGTGKYLNNSPDYLDSPRVGTETNOFRTSLTYPMGCVTLKEEGKRSFNTPTYIRAQ 780
Db 559 FWGTGKYLNNSPDYLDSPRVGTETNOFRTSLTYPMGCVTLKEEGKRSFNTPTYIRAQ 618
QY 781 GDTPESRSIFQSGISDWTYEXIOSVLFPGDINNNLVGVKASSFNSNRPNNGLEMTAA 840
Db 619 GDTPESRSIFQSGISDWTYEXIOSVLFPGDINNNLVGVKASSFNSNRPNNGLEMTAA 678
QY 841 TTYLRSGIGLARTSGLEPNOQFGTTHQVLSVSGOQFSIKNIRTIFFGNOLWYELFTNE 900
Db 679 TTYLRSGIGLARTSGLEPNOQFGTTHQVLSVSGOQFSIKNIRTIFFGNOLWYELFTNE 738
QY 901 NKSQVYTLRLADSSNDPASSSPTSILIDVNEIGVILPLDLSFTYVAAGNVALFSSN 960
Db 739 NKSQVYTLRLADSSNDPASSSPTSILIDVNEIGVILPLDLSFTYVAAGNVALFSSN 798
QY 961 PGSPSYTAVNTFNONISDLAEBSGAKYTSDFMGCTIOFPEDEYLIIONGFTSOVANEVT 1020
Db 799 PGSPSYTAVNTFNONISDLAEBSGAKYTSDFMGCTIOFPEDEYLIIONGFTSOVANEVT 858
QY 1021 NOSFLNSLVDFTPAAGTNRVYVDPDGNLTNOMLPLKVOIQYIDGKYYPAKLNNNLVT 1080
Db 859 NOSFLNSLVDFTPAAGTNRVYVDPDGNLTNOMLPLKVOIQYIDGKYYPAKLNNNLVT 918
QY 1081 FSYNNGALPSWVVPFPAIGSTILGILAMITIIIGLAIIGVPLRAORLQDKGKFTFKKVDL 1140
Db 919 FSYNNGALPSWVVPFPAIGSTILGILAMITIIIGLAIIGVPLRAORLQDKGKFTFKKVDL 978
QY 1141 TAAVGSYKKIITQTNVKKKPPALGAGSGDKRPAAPAKPAPAKSPASBPAPG 1200
Db 979 TAAVGSYKKIITQTNVKKKPPALGAGSGDKRPAAPAKPAPAKSPASBPAPG 1038
QY 1201 PKSGAPTPAPKPAAPKPAAPKE 1224
Db 1039 PKSGAPTPAPKPAAPKPAAPKE 1062

```

RESULT 2

Q9REMY7 PRELIMINARY; PRT; 1062 AA.

AC Q9REMY7
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Putative cytochrome related molecule A.
 GN CRMA.
 OS Mycoplasma gallisepticum.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_Taxid=2096;
 RN (1)
 RC SEQUENCE FROM N.A.
 RP STRAIN=R.
 RX MEDLINE=99061189; PubMed=9846732;
 RA Goh M.S., Gorton T.S., Forsyth M.H., Troy K.E., Geary S.J.;
 RT "Molecular and biochemical analysis of a 105 kDa Mycoplasma
 gallisepticum cytochrome (gapA)."
 RL Microbiology 144:2971-2978(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=R.
 RX PubMed=11083776;
 RA Papazisi L., Troy K.E., Gorton T.S., Liao X., Geary S.J.;
 RT "Analysis of cytochrome-deficient, GapA-negative Mycoplasma
 gallisepticum Strain R."
 RL Infect. Immun. 68:6643-6649(2000).
 DR EMBL, AF214004; AAF25382.1; "
 SQ SEQUENCE 1062 AA; 115803 MW; 7B6D0D9D683A095 CRC64;

Query Match 84.0%; Score 5385; DB 2; Length 1062;

Best Local Similarity 86.0%; Pred. No. 1, 2e-271; Mismatches 8; Indels 162; Gaps 1;

Matches 1053; Conservative 1; Mismatches 8; Indels 162; Gaps 1;

```

QY 1 MNISKKLSYTLIGLAVFGLSASFGFQOSKSDNTOVLNOARTIDANSVRLAGLQ 60
Db 1 MNISKKLSYTLIGLAVFGLSASFGFQOSKSDNTOVLNOARTIDANSVRLAGLQ 60
QY 61 NGSLEPNTLRVDNFTTANGTITIKLSFTKPLYGILSDDCGGYKQVQYSDYTTSMN 120
Db 61 NGSLEPNTLRVDNFTTANGTITIKLSFTKPLYGILSDDCGGYKQVQYSDYTTSMN 120
QY 121 RFDQOTRAYVALLVNDENYHLKRIINTNSRIGNRRNNNSKFEYIGVDNPAHYIRFTDDG 180
Db 121 RFDQOTRAYVALLVNDENYHLKRIINTNSRIGNRRNNNSKFEYIGVDNPAHYIRFTDDG 180
QY 181 TKFNFTNOTGGEIVNDFILDAPIILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSG 240
Db 181 TKFNFTNOTGGEIVNDFILDAPIILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSG 240
QY 241 TNADGMPDGGNGQITNTDPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHKIKT 300
Db 241 TNADGMPDGGNGQITNTDPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHKIKT 300
QY 301 SFOLDERINTNSRIGNRRNNNSKFEYIGVDNPAHYIRFTDDGTFKFNFTNOTGGEIVNDFI 360
Db 301 SFOLDERINTNSRIGNRRNNNSKFEYIGVDNPAHYIRFTDDGTFKFNFTNOTGGEIVNDFI 360
QY 361 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNNADGMPDGGNGQITNT 420
Db 361 LDAPILPKDLHPDWNLVYIQRKILPNDVNTAVVWPVGRVSGTNNADGMPDGGNGQITNT 420
QY 421 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHKIKTSFOLDERFVYPWTGSEE 480
Db 421 DPIAQTCTTTDNONPSTFNSGAMPGANRRYDSQLVKKHKIKTSFOLDERFVYPWTGSEE 480
QY 481 NKNTIRLATGSLPSENERWILDPGTQVTLKEDSVNVFSRLVNSLSFTGSIYIF 540
Db 481 NKNTIRLATGSLPSENERWILDPGTQVTLKEDSVNVFSRLVNSLSFTGSIYIF 540
QY 541 GTSELPSLWYSFPTRLSDLTALNOVKTTDIEASTDNGTTTGTADTSSGSGTACT 600
Db 541 GTSELPSLWYSFPTRLSDLTALNOVKTTDIEASTDNGTTTGTADTSSGSGTACT 600

```

```
|||||
Db 379 GTSLESLWYSEPTRLDLDLALNQYKTDIDEASSPDNCTTNGTTTDTSSGSGAGT 438
QY 601 GNTNTNSQVSNNTLTMTYTSFGIDSKPTSAKIDETNMADPNVTEARITAEYKIGIONET 660
Db 439 GNTNTNSQVSNNTLTMTYTSFGIDSKPTSAKIDETNMADPNVTEARITAEYKIGIONET 498
QY 661 PITNAGNFIRNTIGVGFTSGSRVYLBAVNGDORPTGNFOPLVFGYLGYOQPTGT 720
Db 499 PITNAGNFIRNTIGVGFTSGSRVYLBAVNGDORPTGNFOPLVFGYLGYOQPTGT 558
QY 721 FWYGYKLLNNSPDYLDSPRVGTEINQFRTSLTYPVMGVLTREGARSNTPYTRAQ 780
Db 559 FWYGYKLLNNSPDYLDSPRVGTEINQFRTSLTYPVMGVLTREGARSNTPYTRAQ 618
QY 781 GDPESRSIFOSGYSNTYETIOSVLCFQDIRNNLVAVGKASSFLNSNRNPNGLKIA 840
Db 619 GDPESRSIFOSGYSNTYETIOSVLCFQDIRNNLVAVGKASSFLNSNRNPNGLKIA 678
QY 841 TTYLRSGIIGARTSGLPNOQPGFTHQVLSVSPGDOFSSIKNIRTIPEGNQLWYFLPTNE 900
Db 679 TTYLRSGIIGARTSGLPNOQPGFTHQVLSVSPGDOFSSIKNIRTIPEGNQLWYFLPTNE 738
QY 901 NKSSTYTLRLADSSNPDASSSPSLIDVNEIGVILPLDSEFTVNAAGVAFSSN 960
Db 739 NKSSTYTLRLADSSNPDASSSPSLIDVNEIGVILPLDSEFTVNAAGVAFSSN 798
QY 961 PGSPGSTAVNTNQNLSIDIAFEGSGAKYNSDFMGTOEKFDEYLLONGTSQVARRFVT 1020
Db 799 PGSPGSTAVNTNQNLSIDIAFEGSGAKYNSDFMGTOEKFDEYLLONGTSQVARRFVT 858
QY 1021 NOSFLNSLVDETPANAGTNRVAVDPDGNLTNOLPLKVQIQYLDGKYYPDAKKNLNT 1080
Db 859 NOSFLNSLVDETPANAGTNRVAVDPDGNLTNOLPLKVQIQYLDGKYYPDAKKNLNT 918
QY 1081 FSTNNGALPSSWVPVPAIGSTLGLAIMITLGLAIGIPRAOKRLDQKGFTEKKKVDTL 1140
Db 919 FSTNNGALPSSWVPVPAIGSTLGLAIMITLGLAIGIPRAOKRLDQKGFTEKKKVDTL 978
QY 1141 TAAVGSVYKKIITQTNANVKKKPPALGKSGDKPAAAPAPAPKSPAKSPAKPTG 1200
Db 979 TAAVGSVYKKIITQTNANVKKKPPALGKSGDKPAAAPAPAPKSPAKSPAKPTG 1038
QY 1201 PKSGAPTKPTAPKPAAPKPTAPKE 1224
Db 1039 PKSGAPTKPTAPKPAAPKPTAPKE 1062

RESULT 3
Q9LCS5 PRELIMINARY; PRT; 377 AA.
AC Q9LCS5:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 120-kDa membrane protein MG3 (Fragment).
GN MG3.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S6;
RT Yoshida S.;
RT "Mycoplasma gallisepticum S6-strain gene encoding a 120-kDa membrane protein."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB033310; BAA94277.1; -.
FT NON_TER 377
SQ SEQUENCE 377 AA; 41773 MW; 417B61CDE2B17037 CRC64;

Query Match 27.0%; Score 1733.5; DB 2; Length 377;
Best Local Similarity 65.2%; Pred. No. 1,7e-82;
```

```
Matches 352; Conservative 9; Mismatches 16; Indels 163; Gaps 3;
QY 1 MNISKKLKSTYLLIGLAVFAGLSASGFQSDKSNNTDQVNOARTLDANSVRLAGLG 60
Db 1 MNISKKLKSTYLLIGLAVFAGLSASGFQSDKSNNTDQVNOARTLDANSVRLAGLG 60
QY 61 NGSLEFNTVLRDVDNFTTAAANGTILKIDSTKPLYGDLSDDCGGYVKQIVSDYTTSRN 120
Db 61 NGSLEFNTVLRDVDNFTTAAANGTILKIDSTKPLYGDLSDDCGGYVKQIVSDYTTSRN 120
QY 121 RFDQROTAYALLVNDDEANVHLKRINTNSNRIGNRNNNSKFYIGVDNPAHYIRFTDDG 180
Db 121 RFDQROTAYALLVNDDEANVHLKRINTNSNRIGNRNNNSKFYIGVDNPAHYIRFTDDG 180
QY 181 TKFNTNNOGELVNDPILDAPILPKDHPDWNLTYQKILPNDVNTAVPMPVGVSG 240
Db 181 TKFNTNNOGELVNDPILDAPILPKDHPDWNLTYQKILPNDVNTAVPMPVGVSG 240
QY 241 TNADGMFDCNGQITNTDPIAQTKTNTDQNSFTNSGAMPGANRYDSQLVKKRIKT 300
Db 241 NSATDGTDFRNG-----MPSAN----- 258
QY 301 SFQDERINTNSNRIGNRNNNSKFYIGVDNPAHYIRFTDDGKFNFTNOGELVNDP 360
Db 259 ----- 258
QY 361 LDAPILPKDHPDWNLTYQKILPNDVNTAVPMPVGVSGTNAADGMFDCNGQITNT 420
Db 259 ----- 258
QY 421 DPIAQTKTNTDQNSFTNSGAMPGANRYDSQLVKKRIKTSTFQDEKFEYEMTSEE 480
Db 259 DPIAQTKTNTDQNSFTNSGAMPGANRYDSQLVKKRIKTSTFQDEKFEYEMTSEE 317
QY 481 NKNTIRATGSLPSENERWYLLDIPGTQVTLKEDSVNVEKRLVNSLSFTGDSIYIF 540
Db 318 NKNTIRATGSLPSENERWYLLDIPGTQVTLKEDSVNVEKRLVNSLSFTGDSIYIF 377

RESULT 4
Q9LCS4 PRELIMINARY; PRT; 376 AA.
AC Q9LCS4:
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 120-kDa membrane protein MG3 (Fragment).
GN MG3.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F;
RT Yoshida S.;
RT "Mycoplasma gallisepticum F-strain gene encoding a 120-kDa membrane protein."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB033211; BAA94278.1; -.
FT NON_TER 376
SQ SEQUENCE 376 AA; F99BF61049E0D068 CRC64;

Query Match 25.6%; Score 1643; DB 2; Length 376;
Best Local Similarity 62.2%; Pred. No. 8.6e-78;
Matches 336; Conservative 13; Mismatches 27; Indels 164; Gaps 3;
```



```

121 KVEITOEGBDMSN-----LYLLIPKNTFDPKQASIDPKDLVAYNLTGSSSKQO 172
162 EYI-----GVONPAHVIR--FTDGTKE--NFTNOTGEIYNDPILADPLPK 206
173 ARVVOIENVNLYQSTWTFPTTKGLMHIDPKKMDVNPQMSQSS-----STFETK 227
207 DLHPWVNLXIOKILPN-DVNT-----AVWPW-----VGRVSGNMDDGMEDC 250
228 EDHPSWYVANNSEKXANDADQNTQYNGLSANMVLPMKQYITNLGNPAKNGIYLIIG- 266
251 GNGQITNDPIAQT-----KTTDNQNSTFNSGAMPQA-----NNRYSOLNV 294
287 GNGSIYN-DEALISGMKMLDFELKPSGIDNNQ-----NYGIPAYALLRYLRDPS--- 338
295 KHRKITSFOLDERINTNSNRIGNNNNSKFVIGVDNPAHVIRFTDDGTKEFNFTNQGE 354
339 KPLIGTS-----APNR--RWNOSYAPIGQDNFTYVRLAVGVQJNASTD----- 382
355 IVNDFILDAPILPKDLHPWVNLXI-----ORKLIPDVTAVVWPVWGRVYS 401
383 -----EATYIXLAAGITVQAKESQAREVISNS-NT-----S 413
402 GTNADGMPDQNGQITNDPIAQTCTTNDQNPSTFNSGAMPGANRRYSOLNVKHKRK 461
414 TNKVYTKIDKRSLOLTGANTITNTKDT-----AANSIDPAL-----LFG 453
462 TSFOLDEKFPYEPMTGSEENKNTRLATGSLPSENERWILDIPGTPVLTKEDSVNVSR 521
454 TAFNIDSLINLP---TKLENLTIPO--NVPOYESF--DVGATMSVSAVQYTYEPRK 506
522 -----LYNSVNSLFIQDSIYIFGTSELPWLWYSPPTLSLTALNVQKTDIDIASST 576
507 KMHASSSTTDINTYTTASNGWNNNGRTAF--NSYK-PN--NDIGSFQPKIND----- 555
577 DNGTITNGTTTADTSSGTAGGNTTNTSOVNSNPTLNTY--RSFGIDSKPSTA-NK 632
556 NNNAI-----YSYNLSLIEN-AIVYYSTLSFGYSLCGLGLRK 594
633 IDETNADPNVTEARIVAEYRLGIONEIPITNAGNFIRNTGSGVGFSTGSRVLRASYN 692
595 IEMPSKENP-----ENTTYGAM-QVCKSIYVINEPK 625
693 GQDRPTGNFQPLVYGYLGYOQRTGTFWYG---TYKLLNSPYDLDSPRVGTITNO 748
626 SDRKSTAIHGPSSISIG-----ESNLVGSAKYGDMDYPVYKINNSIGIVPSD--YSNTNN 680
749 FRRTSLTPYMGYLTREGARSPSNT-PYTRAQGTTPESRSIFPGSGYSDN-----TYEYI 802
681 IIMTVAI-----YVT--GIKDFNDTPIIASQFE-----IGNSPYEDNSSTIKTNGTL 727
803 QSVL-----GFGIRNNINLVGVKASS-PLNSRPNPNGLEMAIATYLRSQIGLARTGLPN 858
728 QPTLPWDFVGL-NSINENSEISLMLNNQKTKNNNE-----764
859 QQPFQTHOYIVSPGDQSSIKNITIPGNOLM---YELFTNEN-----KSSYVT 908
765 -----HFIYTKSP--EISE-----YGNAIWTERFYNGSSNNAWKGSRARFE 808
909 LRLADSNPDASSFSPTSLIDNEIGVILPLLDNSFTYV-----N 949
809 VK-----DSNLSNLSSTTVGQVGLDNLADSYVQKNEEQGQDFVLLARTDDT 860
950 AAGNVALFS--SNFGSPG-SYTAVENTENOLSDIAEGSGAKTSPFMGTIOFKPREYLI 1006
861 QKNNDIFEGQJNNRREPISYCKL--KON-----YGSYFETISEIDRLSL 905
1007 ONG-FTSOVARNFTN-----OSFLNSLVDTF--PA 1034
906 GNGQFVNMLSNQILNLANLVOYNLSYTGPNLDSKSTIRIYKNGFLAEVQVQYKNPI 965
1035 NAGTINR-----VVVDPGNLTNOLPLKVOIQIYLDGKYRDAKLKN--NLYTFS-YNN 1085

```

```

966 IEGSTPYGVYIVASNVDFVOSATFTVYSNNSLTKNVDMPRTNSSKINVINNSIFAG 1025
1086 FGLPSWVPVPAIGSTGLIAMIILGLAIGTFLRQORLQDQKFTKFKVDLTAAVG 1145
1026 FSAAMDILPVIAIPYIVLVALIIGLCSIGTGMKAKRAIKVGFELQDVKGLTSAVG 1085
1146 SYKKIIT--TQFANVKKKPPALGAGSGDKRPAAPAKP-AAPAKSPAPRASSPAKPTGK 1202
1086 GVEKKIIDNTSNVSKSPQWL-----KAAAKFPNTVPPARSQLTNDVSRETPPS 1136
1203 SGAPTKPPAPK 1214
1137 S-----APKP 1141

RESULT 8
ID 093006 PRELIMINARY; PRT: 1629 AA.
AC 093006;
DT 01-DEC-2001 (TREMBLrel, 19, Created)
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
DT 01-OCT-2002 (TREMBLrel, 22, Last annotation update)
DE Cyathodesin pl.
GN P.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN 11
RP SEQUENCE FROM N.A.
RC SPMRAIN-MP4817;
RA MEDLINE=21391823; PubMed=11500436;
RA Doi=10.1093/emboj/19.19.5612; Dankert J., Zaat S.A.J.,
RT the pl Cyathodesin gene of individual strains."
RL Infect. Immun. 69:5612-5618(2001).
DR EMBL: AF290000; AAK92038.1;
DR InterPro: IPR004940; Adhesin_P1.
DR Pfam: PF03257; Adhesin_P1.
SQ
SEQUENCE 1629 AA: 175998 MW: C806134A0ABDB6A CRC64:

Query Match 5.0%; Score 322; DB 2; Length 1629;
Best Local Similarity 21.6%; Pred. No. 2.7e-08;
Matches 273; Conservative 141; Mismatches 404; Indels 446; Gaps 70;

251 GNGQIT--NTDPIA-----QTKTTDN-----QNPSTFNSGAMPGANN 286
506 GNGHYTKSAHAPLSIGFVRVYNATGSAVTGWYALLFGSNVKKQTDGLKDPFNNN 565
287 RYDSOLNVKHKRITSFOLDERINTNSNRIGNNNNSKFVIG--VDNPAHVIRFTDD-- 341
566 RW-----FEYPRMAVAGAKFVGR--ELVLAGITTMGDTATVRLLYDEL 608
342 GTRFNETNOTGEIYNDPIL-----DAPL-----LPKDLH--PDW 374
609 ESNLNLVVAQGGIGLREDLQLEFPYGMANRPDLIGAMSSSSSSSSSHNAPYFHNPDW 668
375 YNLXIOKILPN-DVNTAVVWPVWGRVSGTNADD-----GMFDQNGQITN--TDPI 423
669 QDRPIQ-----NVWDFIFPWE-----DKNGKDAAYIYPRYSKM--AMQVYVMSNKL 716
424 AQTCTTD--NONPSTFNS-----GAMP-----GANN-----RYDSOLNVK 457
717 TDQPLASDFVNEAAYOPNSLFAAILPELLAALPDKVYKGENEPAFANETFERFQKLTVA 776
458 -----HRIKTSFO--LDEKFPYEPMTGSEENKNTRLATGSLPSNERY 498
777 PTQGTNWSHSPSLSRSTGFENLVGSVLDQVLDYVWIGNGRYG-----NNHR- 825
499 WILDIPGTPVLTKEDSVNVFSLYLYNSVNSLFT-----GDSIYI 539
826 -----GVDDITAPQISAGSSSGISTNTSGSRFLPTFSGNIGVGIKANYQATLGSQIMI 879
540 FGTSELPWLWYSPPTLSLTALNVQKT-----DIKASSTPDNGTITNGTTADTSSG 594

```

```
Db      880 TCGS-----PRTLDQANL-OLMTGAGWRNMDKSSGSGSDENHFTSATGMDQOG 928
Qy      595 SNGAGNGNTNTNSQ---TVSNPTLNTYRFBGIDSKRTSANKIDETWAD--PVIARIR 649
Db      929 QSGTSAGNDLSLKODKISKSGDLSLTTQ---DGNATGOQ--EATVYTNLPPMLTPT--- 978
Qy      650 AEXRLQONEIPTTNAGN-----FIRNTIGVGFTSGRVYLRAVNDGDPPTNGFOP 703
Db      979 ADM---PALSTFTNNKNNORALFLRGLLGI-----PVLVNSSGSSNKFQATDOK 1027
Qy      704 FLVFGYLGYYQOTRTGFWGYTK-LLN-----NSPYVDLDSPRYG--TE 745
Db      1028 WSYT--DLQSDQTKMLPAYGEVNGLLNPALETVEYGNTRAGSGSNTTSSPGIKRIPE 1085
Qy      746 TNGFRRTSLTYP-----VMGGLTEBARSPSNPTYIRAO----- 780
Db      1086 QNNDKRAKTLTTPCLANFPDQVGNLVVSGTSLFQJLGMV-----SFTDFIKPRAGYL 1138
Qy      781 -----GDTPESRSIFOS-----GYSDNTYEIGSVLGFDGI--RNNLVNGYKAS 823
Db      1139 GJOLSGLDASDSQRELINAKRPMWAFRGSWVRRLGRVESYMDLKGVMDQADQALAAQAT 1198
Qy      824 FLNSN---RPNPGL---EMIAATY--LRQJGLARTS----- 854
Db      1199 SEASGALAPHPMALAFQVSVYBASVSSSTSSSGSGSSNTSPYLHLIKPKKVESTQOL 1258
Qy      855 --GL-----PNO-----QPFCTHOVISVSPDQSSIKNITTTPGNOQLMFTLTNEN 901
Db      1259 DQGLKMLDPNOVTRKLRQSFQDHSF-----QPOSILKTTTPVFG-----TSSG 1302
Qy      902 NKSSVYTLRLADSSNPDASSFPTSLIDYNEI-----GVILPLLDNSEFTVNAAGNALF 957
Db      1303 NIGSVLS-----GGAGGSSSGSGSGSVLSPYERISGMLVGPST-----SDGNTS-- 1350
Qy      958 SSNPGSPGTYAVNTNQ-----NLSDIAEFGSGARYTSDFWGTIQFKPDEYLLIONGFTS 1012
Db      1351 STNNLAPNT---NTGNDVYGVGRLS---ESNAKMDNDVDIVRTPLAELLDEGEQTA 1402
Qy      1013 QVARNVT-----NOSFLMSLYD-TPPAN-----AGTNRV-----Y 1043
Db      1403 DTGPOSVKEKSPQDIDFNRLTTPVTDLPDMYLVYDOYIPLFDIPASVNRKMWRLKY 1462
Qy      1044 VDPDGNLTQNLPLKYOIOLDKYYDAKLNN-----NLVTFSS-----YNN 1085
Db      1463 LSPDNT--BQSLGLRLEFEKPD---QDTQNNNVQVNPNNNGDFLPLLTASSQGPQLFSP 1517
Qy      1086 FGALEPMNVFTALGSLGILAIMIILGLAIGIRLRAQRKLDKGFTTEKKVDTLTAAGV 1145
Db      1518 FNGMPDYVLPLAITVPIVIVLVSTVLGLAIGIRPMHKNKQALKAGFALSNQKVDVLTKAVG 1577
Qy      1146 SVYKKTITQTPANKKKRPAALGAGKSGDKRAAANKRAAPKSPASAPASAKTPGKSSGA 1205
Db      1578 SVKEIINRT--GISQAPKRL-----KOTSAKPGAPRPVPKPGAKPPVQP--- 1624
Qy      1206 PTPK 1209
Db      1625 PKK 1628
```

```
RESULT 9
Q93005
ID 093005 PRELIMINARY: PRT: 1635 AA.
AC 093005:
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Cytadhesin p1.
GN p1.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmales; Mycoplasma.
NCBI_TaxID=2104;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=MAC:
RC MEDLINE=21391823; PubMed=11500436;
RA Dorio-Zetsma J.W., Wilbrink B., Dankert J., Zaai S.A.J.;
RT "Mycoplasma pneumoniae p1 type 1- and type 2-specific sequences within
the p1 Cytadhesin Gene of individual strains."
RL Infect. Immun. 69:5612-5618(2001).
DR EMBL: AF290001; AKK92039.1;
DR InterPro: IPR004940; Adhesin_p1.
DR Pfam: PF03257; Adhesin_p1.
SQ SEQUENCE 1635 AA; 176772 MW; 74FABB8D76091F53 CRC64;

Query Match 4.9%; Score 315.5; DB 2; Length 1635;
Best Local Similarity 20.4%; Pred. No. 6e-08;
Matches 324; Conservative 168; Mismatches 503; Indels 595; Gaps 82;

Qy 41 LVNQARTLDANSYRLAGLGGSLFNTVLRDNDNFTTANGTTIKLDSFTKPLGLDLS 100
Db 219 LPNQGSSSGSDST-----NQTGAMFGLKVD-----ATVDSKOSTESLKGESS 263
Qy 101 DDGGYKVKQIYVDYTSRRRFDOR-----OTRAYYA 132
Db 264 -----SSSTSTSTTQGGSSNENKVKALQAVKSSGSGSGDGTQVE 311
Qy 133 LTVNDEANVLRKINTNSRIGNRNNSKFIQVD--NPAHIVRFTDGTGFNFNTN-Q 188
Db 312 LESNDLANAPIKR-----GSNNQVQLADDFGTAPSSSGSGTODGTPTPTPWL 362
Qy 189 TQGEIYND-----FILAP-----ILPKDL----- 208
Db 363 TTEQIHNDPAKFAASILILYDAPYARRAIDRVHDLDPKAMANYPPSWRTPKWNHGL 422
Qy 209 -----HPDMYN-----LYIORKI--LPPDVNT----- 228
Db 423 WDKRADVLIQTGFENPRRHPFEDGGQTVADNEKTEFDVDNSENTKQFQKEADSDKS 482
Qy 229 AVVPMP-----VGRVSGTNADGMEFCNGQIT--NTDPIA-----QTKTT 267
Db 483 APIALPEAFANIGNLTWFGQALLVFG--GNGVHTSAHAPILSIGFRYRNATGTSAT 541
Qy 268 TDQNPSTNSGMPGANNRYDSQLNVKRIKTSFQDERINTNSNRIGNRNNSKVIIG 327
Db 542 VTGMPYALLFSGAV---NKQTDGLKLNLFENNRMFEVPPMAVAKFVGR---ELVLA 594
Qy 328 G---VONPAHIVRFTD--GTFENFTNQTGELVNDIL-----DAPT----- 365
Db 595 GTTMDGTATVPRLTDELESNLNLAQGGGLREDLQLTFTPGMANRPDLPGANSSSS 654
Qy 366 -----LPKDLH--PDWYNTLYORKILPNDVNTAVVPVGRVSGTNAD-----G 408
Db 655 SSSHNAPYFHNHPDMQDRPIQ-----SVVDARLIKWE---DKNGDKAKYIYPYRSG 705
Qy 409 MEDCGNQITN--TDPIAOKTKTTTD--NONPSTNS-----GAMP-----GANN 448
Db 706 MW---AQOYVNMENKLTDPDPLADFEVNEAYQPSNLSFAAILNPELLAALDKKYKGENE 762
Qy 449 -----RYDSQNLNVK-----HRTKTSPO-----LDEKFVYPEWMSSENK 482
Db 763 PAANEYERENOKLTVAPTOGTNMSHSPTLSRSTGFLNGLVSDVLDQVYPIWIGNGYR 822
Qy 483 NITRLATGSLPSNERWIIDIPCTPOVTLKEDSVNFSRLYLNSVLSPTI----- 533
Db 823 G-----NNHR-----GVDDITATQAGSSSGISTMTSGSRSLPFFSNIGVG 865
Qy 534 -----GDSITYIFGISELPSLWYSPPTPLSLDITLALNOYKTD---DIFASSTNG 579
Db 866 LKANVOATLGGSGTMTTGGG-----PRTLDQANL-OLMTGAGWRNMDKSSGSD 914
Qy 580 TTNNGTTTADTSSGSGAGTGN-----TTNSQVYVSNPTLNTYRSFGI 623
Db 915 DHTKFTSATGMDQGGSGISAGNPDSLKODKISKSGDSLTTQDGNATGQOEATVYTNLPP 974
Qy 624 DSKPT-----SANKIDETNW----- 638
```

[illegible][illegible]

```
Db 1094 TSKAVLITPGLAMTFQDVGNLVSGTSFSGFOLGMLVFTFDFIKPRAGYLIGLQTLGLDAS 1153
QY 715 -QTRGTFWYGYTKLLNSPVDYLDSPRY-----GTETQOFRTSLTTYVMGYLTLEEA 768
Db 1154 DATQALIV-----APRPMAPRGSMVNRILGRVESVMDLKGMADQ--A 1195
QY 769 RESFSTPYIRAGODT--PESRSIFQSGYSDNTYEIOSVGFQIRNNLVNG----VKASS 823
Db 1196 QSDSGOSTTATRDALPE-----HFNALAFQVSVEASAKYKNTSSGQTOSTNSSP 1246
QY 824 FLNSNRPN-----PNGLEMIATYLRQSIGLARTSGLPNOQPFQTHOIVSYSPGD 875
Db 1247 YLHLVKKPKVIOQSKLDLDDKILDPNQVTKL-----RQSGFGTDH---STQPPQ 1293
QY 876 QSSSIKNITITFPQNLWYFLFTNNENKSSVYTLRLADSSNDPASSFSEPTSLIDVNEI- 934
Db 1294 Q--SLKTTPTIFG-----TSSGNLSSVLS---GGAGGSSSGSGQSGVDLSPYE 1337
QY 935 ---GVILPLDINSFYTVNAGNVALFSSNPGSPGSAVTAVTFNQ-----NLSDIAPEGSG 986
Db 1338 KYSGLWVQLPST-----SDGNTS--STNNLAPNT---NTGNDVYGVGRLS---ESNA 1382
QY 987 AAYTSDFMGTIOFKPDEYLIONGFTSQVARNFVT-----NOSFLNSLVQ--FTYRAN- 1035
Db 1383 AKMNDVDGIVRTPLAELLDEGEGQFADTGPQSVKFKSPQIDFNRLFTHPVMDLPDPVTM 1442
QY 1036 -----ACTNRYR---VYDPDGNLTNQNPLKVOIQYLDKKYDADAKKN- 1076
Db 1443 LVYDQYIPLFIDIPASVNPKNWRLKYLSPDTN--EOSTGLRLEFFKPD---ODTOPNNV 1497
QY 1077 -----NLVTES-----YNNFGALPSWVVPYTAIGSTGLIAMIILGLAIGPL 1119
Db 1498 QVNPNNGDLPPLTLASSQGPQLFSPFNQMPDYVLPALTIVPIVLSVTLGLAIGTM 1557
QY 1120 RAQRKQDKGKFTTEKKVDTLTAANGSVYKIIITOTANYKKRPALGACKSGDKRPAAA 1179
Db 1558 HKNQKQALKAGFALSNCKVYVLTAKVGSVEKEIINRT--GISQAPKRL-----KOTSAA 1608
QY 1180 KPAAPAKPSAPKASSPAKPTGPKSGAPTKP 1209
Db 1609 KKGAPRPVYPKKGAPKPPVQP---PKRP 1634

RESULT 11
Q8ID65 PRELIMINARY; PRT; 1813 AA.
AC Q8ID65;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF13.0339.
OS Plasmidium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Batteelli B.,
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844509; CAD52762.1; -.
DR Hypothetical protein.
SO SEQUENCE 1813 AA; 7EDA3299CD6634DC CMC64;

Query Match 4.0%; Score 259.5; DB 5; Length 1813;
Best Local Similarity 18.6%; Pred. No. 5.6e-05;
Matches 191; Conservative 143; Mismatches 335; Indels 357; Caps 44;
```

```
QY 165 GGVDNPAHVIRFTDQTKFNFNQGEIYVDFILDAPILPKDLHPDWNYLQKILPN 224
Db 1099 -----NNIN-----ITNN-----NNIINNIIISH 1117
QY 225 DVNTAVVPWPVGR-----NNIN-----VSGTNADGMEGCGNGQITNDPIA 262
Db 1118 NMNTPVANKITKSKEDNEYLISLRASDQCKRSIIENHNDEYFPLNKSNSFNLQHP 1177
QY 263 QRTKTTDQNPSTFNSGAMPANR-----YDSQLVKHKRITSQDLERITN-- 311
Db 1178 DSITNNMCN---TYAICYEGAINNOGIIIGIKPYDN--INKINNVPTEFNMSSALNTNNI 1232
QY 312 -----SNRIGRRNNSKSFVIGGVNDPAHVIRFTDQTKFNFNQGEIYVDFILDAP 364
Db 1233 FNTNALTANNIFKKN---VFG-----TFTSTNLGCTN----- 1263
QY 365 ILPKDLHPDWNYLQKILPNDVNTAVVPWPVGRVSGTNADGMEGCGNGQITNDPIA 424
Db 1264 -----NSTLNPIPAISTGCT--FNSNIFSSGTSNIFGTNNMT 1298
QY 425 QTKT-----TTDQNPSTFNSGAMPANRYSQLVKHKRITSQDLER 469
Db 1299 STNVFNNTNNVSTNIFGTTNTQPSMFTG---GTNNSL-----IGTTNTSF----- 1345
QY 470 FYVPEWTGSEENKN--ITRLATGSLPENERWILDIGTPQVLTKEDSVVFESRLYSVN 528
Db 1346 -----SALNKNKTLNNSTAN-----NMNTNMNNTLNNIIIS 1378
QY 529 SLSEIGDSIYIFGTSELPSLWYSPFTRLSDLTALNOVKTDDIEASTDN--GTTNGT 586
Db 1379 SSSNIFNKDKLFGSSGTPNIF-----NNNTLSNTANTFGSTNTS 1420
QY 587 -----TTADTSSGSGTGCTGNTTTSQTVSNPLNTYRSQIDSKPTSAKIDETNADR 642
Db 1421 IFSKNJANSTSGNLFEGNTGTTNMENTSNNNLFN-----NNNNNNNNNNNN 1468
QY 643 VIEARIYAER--IGIONEPIITNAGNFRNTIGVGF-----TSTGSRYLRSAYGDR 696
Db 1469 SPFNSLSYSSNNNSLNLSNLSL--INNNTN--NMGSSGFNRNRSTNSLGLTITTTNNNN 1525
QY 697 PTGNFQPLVFGYLGQOTRGTFWYGYTKLLNSPVDYLDSPRVGTETNOFRRTSLTY 756
Db 1526 MTNNI-----LNKSSQNVSSNLFGNH--NNSNNIFSSMNTSOGNKI----- 1566
QY 757 PYWGYLTDEGARSPFNTPIYIRAGDTPESRSIFQSGYSDNTYEIOSVGFQIRNNIN 816
Db 1567 -----NTYDKKNLFSSNLNNNTSSMLKD--NNEFASWGNKN 1600
QY 817 VGVKASSFLNSNRPNNGLEMIATYLRQSIGL-----ARTSGLPNOQPFQTHOIVSY 872
Db 1601 QMNNNSFYLVN-----LNTSIFGSKKREINONNNSNLTFLNNSG-----YIGV- 1646
QY 873 PGDSFSIKNITITFPQNLWYFLFTNNENKSSVYTLRLADSSNDPASSFSEPTSLIDV 932
Db 1647 -GNNTYANI-----IGNNN-----SNANNN-----NNVSSSSSSLFSSNNMFNTN 1685
QY 933 E-----IGVILPLDINSFYTVNAGNVALFSSNPGSPGSAVTAVTFNQNLSDIAPEGSG 988
Db 1686 SNNMFGSKNGLLGTS--NANNVSNLLLNNSNANNIFN--NKLPTNTNNNFTSGTNTN 1741
QY 989 YTSDFMGTIOFKPDEYLIONGFTSQVARNFVNTQSPFLNSVDETPANAGTNRVVVYDPPD 1048
Db 1742 ITKNTFSVNTNMNMNNNIPQN-----NNSVSVKN--ISQISNNTNNTSSNNNNNNIIQOKN 1793
QY 1049 NLNTON 1054
Db 1794 TTSNFN 1799

RESULT 12
Q8ICT8 PRELIMINARY; PRT; 2752 AA.
ID Q8ICT8
```

AC 081CT8: 081CT8: 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Hypothetical protein.
 OS MAL6P1.80.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxId=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
 RA Bertrian M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 KW EMBL: AL844505; CAD50351.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 2752 AA; 325015 MW; 44539CDE80FF64F CRC64;

Query Match 4.08; Score 257; DB 5; Length 2752;
 Best Local Similarity 19.38; Pred. No. 0.00013;
 Matches 255; Conservative 165; Mismatches 508; Indels 390; Gaps 57;

OY 3 ISKRL---KSYTLIGLAIVGALGASFGFGKSDKSDNDTQLVNQAFLDANSVRLAGLG 59
 DB 1553 INKNIQHNKNYDIINNNIFONRGSNNIGSYDDRNQKQLOFNTLYHID--VALSPV- 1608
 OY 60 QNGSLFNTVLVDVDFNFI---TAANGTIKLSDFKPYGLDSDCGGKYKQIVSDYT 116
 DB 1609 -WTKRIKNI-KDIFNMKIKKNCNDIINNDDILKNDHDEFINDO-QHTNINVAHFEKR 1665
 OY 117 TSNRFDQDQRA---YYALL-----VDEANVH----- 142
 DB 1666 TLNRYKQDNTGATNVVSPILNITTLKTKKIKKKKYESFPDINDVDVHNNKTEON 1725
 OY 143 -----LKRINTSNRIGNRNNNSKVIJG----- 166
 DB 1726 KDNQHIENNINNINNINNINNINNINNINNINNINNINNINNINNINNINNINN 1785
 OY 167 ---VDNPAHVIRFTDDGKFNFTQGEIYNDILDAPIPLKDLR-DWYLYLQKTI 221
 DB 1786 YIIPEDNPIH-NNFTYDNKKTKTKTQTLQHLNTEINFLP--DIDPIDVTNINIKR- 1839
 OY 222 LPNDVTAVVPMVPGVSGTNADGMEFCGN-----GO 254
 DB 1840 --NKNMKIYTKVYNHSHHTNKSMEFEPNNKLDSTEDRNIOFYENGNRYMNDYKQ 1897
 OY 255 ITNNDPIAOKT-----TTDNQNPSTENGAMP-----GANNRYDSOL 292
 DB 1898 --STPLISGTTNNIRKGTGESTPLISGTTNTTRKGTGS--TLPLISGTTNNIRKTER 1954
 OY 293 NVKRIKTSFQIDE-RINTNSNRIGNRNNNSKVIJGVDPAPVIR-----FTDDGK 344
 DB 1955 FPKAKNSQENINONEINYNNDNDDDDNNNNDDDDNNNNMLNNOYGDIMTESNYI 2014
 OY 345 ENFTQOGEIYNDILDAPIPLKDLRHWNYLYIQKILPNDVTAVVPMVPGVSGT- 403
 DB 2015 SSHNNLPQEDIDNDEKTIEMIEKTIIDQYIN--IEDK---NPKHDMIKIKRKNNNI 2069
 OY 404 -NADGMEFCGNQITNTDPIAOKTITTDNQPSTENGAMPANRFD-----SOLNVKH 458
 DB 2070 HYNNDNNND--NNNDNNND--NNNDNNNDNNNDKFIED--KQINNVDENNINNNIHLKD 2123
 OY 459 RIKTSFQIDEKFEYPMGTSEENKNITRLATGSLPSNERYILDIPTQVTLKEDSVN 518
 DB 2124 NIK---KKKEKNENYKISSYNNN-----DQDN- 2149
 OY 519 FSRLYINVSLSFIDGSIYIGTSELPSIAY-----YSF 553
 DB 2150 ---YGDNNNNL--INNSIYYDSKRSIKIGDEYEETVTKVNIQKNYGININNDY 2203
 OY 554 PTRISDLALN-----QVKTDLIASSTDNGTGTGTTGTTADTSSSGTAGGNTNT 606

DB 2204 NVIFTNIAQVNDENLYFNRSRINIDINDDNNCKVKNITYINNNSNPLESNNNGSTTTT 2263
 OY 607 SOTVSNPTLNTYRSFGIDSKPTSAKIDETWMAHPNVIETARIYAEYRLGIONEPIPTNAG 666
 DB 2264 TTTTS--IKNNKHVNVD-----INNLDI-----KRIYVNSN 2295
 OY 667 NFRNTIGVGFTSGSRVLYLASYNGDQRLGNEFQPLVYEGYLGYYQQRTRGTGFWGY 726
 DB 2296 N-----NDSTVNERDNNININEHGE-KPCY---YTHYNNVNSNININFR 2338
 OY 727 KILNNSPYDLDSPRVGTETNOFRRTSLTPYMGVILEGARSFNTPYIRAGDPDES 786
 DB 2339 --NPSLNLNPKKNNQDSI-----FKKQYSSYPVAKND 2374
 OY 787 RSIFQSGYDNTEYIOS-----VLGFDGIRNLNVGK-----ASSFTN----- 826
 DB 2375 INIFYSKYDNSNRYIEOKIKKNENDVLYKNDKIFSLNKSSECVHSSNSFNANLYIP 2434
 OY 827 -SNRPNPGLEMIATYTLNSQGLARTSGLPNQQFGTTHOVIVSPQDQFS---SIK 881
 DB 2435 YNNMKDKNIMSPONNIIYDEQIINNINFTL--NNHNNNGHNT-----GDEYSDSEYER 2488
 OY 882 NIRTIFPQNOJL---YELFTNENKSSVYTLRLADSSNPDASSSPSLIDVMEIGYI 937
 DB 2489 HINNTINKDNLDEMNNIINNNDNNNNNDKSNLYLHNNNDKSNHFLHNNNDSTVYDIY 2548
 OY 938 LPLDNSFYTVNAGNVALFSSNPGSGSYTAV--NFTNOMLSDIAE----- 983
 DB 2549 ---NHFTYV---NCLNYDNDKNELSKCAVLRKNEQONPIREHEYNISYPSLIK 2600
 OY 984 ---GSAVYTDQFQTOFQFPEYLLQNGFTSOVARNFYVNSQFLSL--VQFPA--- 1034
 DB 2601 NKNNSKTKYLDNLSTLNKDEKNYHPEKKNFNNEKN--NKIFTDIEHISEFNKKICK 2657
 OY 1035 ---NAGTYRVVDPDGNLNONILPKVOIQYIDGKYUAK--LKNNNLVTFSTYN 1085
 DB 2658 IKNIYSKNNKQFFKNNTHLFTNTINLYEKKNYVDDEFSHKQDRIKSSLIKRNKN 2715

RESULT 13
 O8CP76 PRELIMINARY; PRT; 9439 AA.
 ID O8CP76;
 AC 08CP76;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE dbha protein.
 OS Staphylococcus epidermidis.
 OS Bacteria; Firmicutes; Bacillales; Staphylococcus.
 OC NCBI_TaxId=1282;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SPRAIN-ATCC 12228;
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Ju Y., Qin Z.,
 RA Chen Z., Wen Y.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 RW EMBL: AE016747; AA004725.1; -
 KW Complete proteome.
 SQ SEQUENCE 9439 AA; 1050771 MW; 5A8C68E9AA31CDD4 CRC64;

Query Match 3.98; Score 252; DB 16; Length 9439;
 Best Local Similarity 20.48; Pred. No. 0.0012;
 Matches 284; Conservative 148; Mismatches 407; Indels 556; Gaps 73;
 OY 31 OSDKSNQNTQLVNQAFLDANSVRLAGLQNGSLFNITYLRVDQDFTIANGTIKIDSF 90
 DB 634 QATKANSNR-----SNINSEKQLAERSNGYSKSIIRDDEKSYLANSN----- 679
 OY 91 TKPLVGLDSDCGGKYKQIVSDYTSRRFRDQROTAVYALLVDEANVHLKINTNS 150
 DB 680 --PIQVIDLVBPDMGCGGROY-----SHS 701

QY	151	NRIGNRNNNSFEVIGVDNPAHVIREF--DOSTKRNPNQ--TQSGIVNDELIDAPILPKDL	208
Db	702	NVITNENKS--TVNGQVPEANGASAFENIDKVKKMAANNIGVYYAQQILYADLPSPRG-	759
QY	209	HPDMVNLIIQIOT-----KILPND-VNTAVVPMPEVGR-----VSG-----T	241
Db	760	-----YIEELGONLSNTNNVIVVYVPSDKVNPSTI--VGNVHHHTYSGTEFKNTI	809
QY	242	NADDMGFDCGNGQITNTDPLAQTKYTTDN-----ONPSTENS-----GAMPANNRYD	289
Db	810	NVND---NYGLNTPVASTSDSAITGTRNNNELVGAQPVNTSINKIVKATDKSGNESIV	866
QY	290	S-----QLANKHKIKTSFOLDEKININNSNIGRRNNNSKRVIGCVGNPAHVIFETDD-G	342
Db	867	SEYVNIKPIPLNKRYKITTS-----NOTPRAVISIQNNANLISIDONRVSSLSMKTKLG	921
QY	343	TKENFTQTOGEIYNDELIDAPILPKDLHPDMVNLIIQRIKLPRDVTAVVPMPEVGRSG	402
Db	922	TR-NYVESN-----NDVMSQVYS-KVNR-SG	945
QY	403	TNADDMGFDCGNGQITNT-----DPLAQTKYTTDNONPSTENSFGAMPANNRYDS	452
Db	946	NNATVNVNTTSPSDGTWTITVPRVHVLLLEVPTRTYRGQQ---FPGKGTSPNDFES-	1001
QY	453	QLANKHRIKTSFOLDEKFPVPEMTEGSEEN-----KNITRLA	488
Db	1002	-----LRITGGPVDAIRVWVNNQGPDIRNSIQIGRDLTLHAEIFEDGETTPIRKDTYKXL	1054
QY	489	TGSLPSN-----ERYW--ILDI-----PGRPQ-----	508
Db	1055	SQSIRKQIYETTINGRFRSSGDALPGNFQVAVNQYWPBHMFRMAQSGSITPSSKNAGSFT	1114
QY	509	--VTL-----KEDSVNPFSLYLS--VNSLSF-----GDSIYIFGTSELPSLWYYSF	553
Db	1115	KTVVYVYVQNGQTEKENVNLFKVKPKPKPYIDSNSVTSKGLNQOQILVRVWPNQAAVTLXQ-	1173
QY	554	PTRLSDLLALNQVATDIEAASSDNGTNTNTTADSSGST-----GACTGVTNTMSQ	608
Db	1174	-----SNGVITPNTNTTIDNSGIAVTYTOGTLPGNITAKTS	1210
QY	609	TVSNPPLTYRFSFGIDSKPT-----SANKIDET-----NMADPN	642
Db	1211	MTNNVYTYKONSSGIASTTTEDISVFSNSDQVAVTAGMAKNGIKIKNTNNNFENDN	1270
QY	643	VIEARIYAEVRLGION-----IPIT-----MAGNF	668
Db	1271	SFISNIPASHSTLTWNEEPNSKNNIGTTKTIVTLPNHOQTRVYDIPITITPYMTYAKNP	1330
QY	669	IR-----NTIGCVFTSGSRVVLKASVNGDQRTGNF--QP	703
Db	1331	VRDQGRNLGTGVNYIIFENNNRLGCT-----ASMKNDRODKPIAGVON	1378
QY	704	FLVYFGYLG-----YQOTRTGTF--WYGYKLLNS--PY	734
Db	1379	LIALVNYGISTPLEVPKVVVNPFDQPIYIKIQVGDTEFPKGIWAGYIKLENGECLPI	1438
QY	735	D-----VIDSPRVGETNQ-----FRRTSL--TYPV-----GGYLTEGAR-----SF	771
Db	1439	DGMKFYVWQSGTGTSDQMOQLAVTRPFVKYTGTYDVVNPMSMGVWQSGAKFIYTNAK	1498
QY	772	SNTFYIRAO--GD---RPES-RSIFQSGYSNTEYIOSVYLFGDGINNNLVGKASSEL	825
Db	1499	NPOPIITOSKGDVYVPGAVRNLIISG---TNDYIDA--SADKIYINKN-GNKLTFV	1551
QY	826	NSN-----RPNPNGLMIATFYRSQIGARTSGRLNQOPFGTHOIVISVSPDO	876
Db	1552	KNNDGRMTVEPSPINCI-----GPTNNGTAISLR-----LAVRPD-	1590
QY	877	FSISIKINRTIPPGNOLWYFETJNENKNSVYTLRLADSSN-----PDAS-S	921
Db	1591	--SIEALITBSGERTI-----SWSATSEIITVAVAPQEOVATHTHYDNGTDLIPDNERN	1642
QY	922	SFSPTSLDIVNEIGIPLDLNSEFYTVAAAGNVA--LESSNPGSPGTYAVN---TFN	974

Dd		1643 SLNTERVEIN-----YTEKLNENEFQKSTFTTKNNNGKTTINKKNYEEN	1689
Oy		975 ONLSDIAFEESGAKYTSDFMGTIOFKPDEYLIONGFTSOVARNFVNTOSFLNSLVDETPA	1034
Dd		1690 QDNCKVVFSAATIKPNISQI--TIRPKAGQGCTEN-----TNPTVIQA-----PA	1721
Oy		1035 NAGTNRKYVVDPDGC-NLTNQMLPLKVQIOYIDGKYYDAKLKNNNLVTFFSYNFCALPSWV	1093
Dd		1732 QHTLTINEIVKEEQONVTINDINNNAVQ-----PNKNRVALKGN--ALPP--	1775
Oy		1094 VPTAIGSTGLITAIM 1108	
Dd		1776 -NLGGSTSHIPVI 1789	
RESULT 14			
ID	O8GR92	PRELIMINARY;	PRT; 4727 AA.
AC	O8GR92:		
DT	01-MAR-2003 (TREMBLrel. 23. Created)		
DT	01-MAR-2003 (TREMBLrel. 23. Last sequence update)		
DT	01-MAR-2003 (TREMBLrel. 23. Last annotation update)		
DE	GI1521.		
GN	GI1521.		
OS	Mycoplasma mobile.		
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.		
OX	NCB1_TaxID=2118;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RA	Seto S., Uenoyama A., Miyata M.;		
RT	"Identification of Glis21, Force Generating Protein, for Gliding		
RL	Motility of Mycoplasma mobile."		
SL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB084761; BAC23069.1; --		
SO	SEQUENCE 4727 AA; 520559 MW; 24BEC67133DBD2EC CRC64;		
Query Match 3.9%; Score 250.5; DB 2; Length 4727;			
Best Local Similarity 20.8%; Pred. NO. 0.00059;			
Matches 282; Conservative 176; Mismatches 487; Indels 409; Gaps 71;			
Oy		19 FGALGSASFGKOSDKSNDNTOLYNQARTLDANSVRLAGL----QNGSLFNTVLRDVD	73
Dd		1622 FETLKSTSVIIGSSDSKN---LSINFQLSLSSLRSGSIDVOFSINNGTWESYL--VE	1675
Oy		74 DNFIITA---NGTIIRKIDSFPKRLPGDLSDDDCGYKKQOIVSDYTTSRNFDQRQPRAY	130
Dd		1676 GOLIDASVSNDPTKYSRLRPFRRNPFAEISIQ-----	1707
Oy		131 YALLV-NEAVVHLKRINTNSNR--IGNRNNSKFVIGVDNPAHVIRFTDDGKFNFT	186
Dd		1708 -ALVVKYNAQAILISGLININLRDANLVNLSGNTRDLAINDO--NVYRGATGNSPTPI	1764
Oy		187 NOTOGELVDFILDAPIL.PKDLPDWNYLIQRKILPDNDVTPAVVPWVGSGTNADG	246
Dd		1765 NIIVORFAINGVLGGTFPEEKD-----EFLRELNSLPKSIIINAIPRCDF--HVSFILLNNE	1817
Oy		247 MEDCGNG-----QINTTDPIAQTKTTDDQNOSTFNSGAMPANNRYD--SOLNVKHRITS	301
Dd		1818 DFLSEEGISTPOIPITNVKKRLVLDLHDVNLANGSISSGSTSLSLSNPSSLSETTL	1877
Oy		302 FOLDERINTNSNRICNRNNSKFVIGVDN---PAHYVIRFTDDGKFNFTNOTOGELVND	358
Dd		1878 TOLGIVLEKGYSTR--RANGKIDVINYWEHTIIPAISAIRPDRINQNF-----	1925
Oy		359 FILDAPIL.PKDLPDWNYLIQRKILPDNDVTPAVVPWVGSGTG-----ADDGMEDCG	413
Dd		1926 -----PAPINSNASFSLAFREKII--DEMTTII-----VNGSNSINVNCGDV--DLG	1968
Oy		414 NGQIINT---PPIAQTKTTDDQNOSTFNSGAMPA-----NNRYDS--QLNV	456
Dd		1969 PSINNTSLKRIIFCTTTTSMFINNDLOKTVLGTTTININOLISHSENLLNSTIIGTDI	2028

NCBI:taxid:7074

SEQUENCE FROM N.A. cypnus;
STRAIN=Kenya tick cypnus;
MEDLINE=93194085; PubMed=7680636;
Gillmore R.D. Jr.;
*Comparison of the rompa gene repeat regions of Rickettsiae reveals
species-specific arrangements of individual repeating units.*;
RI Gene 125:97-102(1993).
RL EMBL, L01462; AAA9908.1; .
DR NON_TER
ET NON_TER
SQ SEQUENCE 1029 AA; 99999 MW; 09573881A5B9BDCL CRC64;

Query Match 3.9%; Score 247.5; DB 2; Length 1029;
Best Local Similarity 20.9%; Pred. No. 0.00011;
Matches 269; Conservative 131; Mismatches 468; Indels 417; Gaps 62;

7 LKSYLLIGGLAVF-GALGSASFGFSKDSKSNQTQLVQACATLDANSYRLAGLGQNSLF 43
7 LRTVAVGAGITTEGAIKATV-----TTKLTNAASVLTLTNV----- 43
66 NTVLEDDVDNNTTAANGTIILKDSFTKPLVG-----LDSDDCGYKVKQIVSDYTT 117
44 NNVLTGALIDNTTGVNDGVLTNLNGALSQVTGNGIGNATLATTISVGAGATGAGAVIKAT 103
118 SRRFDQROTFRAYALLY-----NDEAN---VHLKRIPTNSNRIGNRN-----NSK 161
104 TKLTLDASQVTFEFTNPVVVTGALIDNTGANNNGCIYTFEGDSYVGINIGNATLATTISVGAG 163
162 FVLGVDPNPAHVIREFTDGTFKFNFTNO--TQGEIVNDFILDAPILPKDLHPDWNLYIQR 219
164 ATLGALIKATTTKLTLDASQVTFEFTNPVVVTGALDN----- 199
220 KILPDNVNPAVWP-----VGRVSGTNADDMGFCNGQIINTPDIATQKTT--TDNON 272
200 ---TGANNNGCIYTFEGDSYVGINIGNATLATTISVGAGATGAGAIKATTTKLTLDAS 255
273 PSTF-NSGAMPGANNRYSQLVKHKIKTSFQDLERITNSNRIGNRN-----NSKF 324
256 QVTFEFTNPVVVTGALDN-----IDNTGANNNGCIYTFEGDSYVGINIGNATLATTISVGAG 308
325 VILGVDPNPAHVIREFTDGTFKFNFTNO--TQGEIVNDFILDAPILPKDLHPDWNLYIQR 382
309 TLGGAIKATTTKLTLDASQVTFEFTNPVVVTGALDN----- 343
383 ILPDNVNPAVWP-----VGRVSGTNADDMGFCNGQIINTPDIATQKTT--TDNONP 435
344 ---TGANNNGCIYTFEGDSYVGINIGNATLATTISVGAGATGAGAIKATTTKLTLDASQ 400
436 STF-NSGAMPGANNRYSQLVKHKIKTSFQDLERIVPEWGTSE-----ENKNTIRLATG 490
401 VFTFNPVVVTGALDN-----IDNTGANNNGCIYTFEGDSYVGINIGNATLATTISVGAG 445
491 SLPNSRERYILDIPTGPVLTLEDSVNVFSR-----LYNSV-----NSLSF 532
446 NVAG-----LKVQGG---VKSNTINLTLDASQVTFEFTNPVVVTGALIDNTGANNNGCIYTF 498
533 IGDSTIYI-----FQSELPISLMYSPFTRLSD-----LFLANQYKTDDEA 573
499 TGDSTIYVGINIGNATLATTISVGAGATGAGAIKATTTKLTLDASQVTFEFTNPV-----VYT 554
574 SSTDN-GTTTNGTTTADTSSGTSIGAGCINTNTNSQVSNFTLNTVYSPGIDSKPTSAK 632
555 GAIIDNTGANNNGCIYTFEGDSYVGINIGNATLATTISVGAGATGAGATGAGATGAGATGAG 610
633 IDEINWADPNVIERIYAERLGIQNEIPITNAGNFINTTIGGVGTFISTGSRYLVRASYN 692
611 TDNAS-----QVTFNPVVVTGALIDNTGANNNGCIYTFEGDSYVGINIGNATLATTISVGAG 643
693 GDSRSTGNFQFLVYVGLGYQOTRGCTFWYGYKLLNNSPYDLDSPRGVSTETNOGRRRT 752
644 GDSYVGTG-----IG-NTNAL----- 656


```

QY 753 SLTPVWGVGLTEEGCASFSTPEYIRAOQDPE---SRSLFQSGSDNTPEYIGSYLVGFD 805
Db 659 -ATYVWAGGLLRQGGVYKSNY--INLTDNASQYTFNPPVYVYGAIDNT----- 704
QY 810 GIRNNLVGVKASSFLNSNRPNGLEMI--AATYLRSQIGIARTSGLPNOQPFQTH 866
Db 705 GNANGNIVTEFTGNSTYGNIGNTALATVWVGAGIATLEGAVIKATTTKLTILNASVLTLL 764
QY 867 QVIVSPG--DQFSIKINRTIFFGNQLWTFLEFNENKSKSVYTLRLDSSNPASSSFS 924
Db 765 NVNVLVGLAIDNTTGVNNGVL-----NLGALSQYV--GNIGNTALATIS 809
QY 925 -----PTSLLD-----VNEIGVILPLDSEYTFVAGNVALF---S 958
Db 810 VGAGKATLGGAVIKATTTKLTLDNASQYTFNPPV-VYVIGALDN---TGNANGNITFTGDS 865
QY 959 SNPSGSPSYRAVNTFNONLSDIAEFGSGAKYTSDFWCTIOFKPDEYILIONGFTSQVARNF 1018
Db 866 TVTGNIGNTALATV--VGAGL-----LRVQGGVYKSNYINL 901
QY 1019 VTNOSFNSLVDTFPANAAGTVRYVYDPDGNLTQNLPLKQIQYLDGKYYDADKLKNNLL 1078
Db 902 TDNA-----SAVTFNPPVYVIG--AIDNTG-----ANNGI 930
QY 1079 VTESYNNFGALPSWVPEPAIGSTGLIAMIIIGLAI-GIPLRA-ORKLODKGFKTFK 1135
Db 931 VTFEGDS-----TVTGNIGNTALATISVGAGKATLGGAIKATTTKLTLDNASQYTF 983
QY 1136 KVDPLTAAY--GSYVKIILITQTN 1157
Db 984 NPVVYVYGAIDNTGNANGNIVTEFGD 1008

```

Search completed: October 10, 2003, 16:35:18
Job time : 89 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 13:34:27 ; Search time 21 Seconds

(without alignments)
2740.985 Million cell updates/sec

Title: US-09-901-572A-4

Perfect score: 64.3
Sequence: 1 MNISKKSLKSYTLIGSLAVFG.....APTPTAPKAPAPKPTAPKE 1224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598.5	9.3	1052	1 MGPC_MYCGE	P22747 mycoplasma
2	556.5	8.7	1218	1 MGPC_MYCPN	Q50341 mycoplasma
3	501.5	7.8	1122	1 ADPL_MYCGA	Q49379 mycoplasma
4	328	5.1	1627	1 ADPL_MYCPN	P11311 mycoplasma
5	285	4.4	1444	1 ADPL_MYCGE	P20796 mycoplasma
6	232.5	3.6	1021	1 OMPL_RICCN	Q52657 rickettsia
7	216.5	3.4	1643	1 OMPL_RICPN	Q53020 r outer mem
8	215	3.4	1656	1 OMPL_RICPN	Q06653 r outer mem
9	204	3.2	1655	1 OMPL_RICCN	Q9Kka3 r outer mem
10	202	3.1	1378	1 YS89_CAEEL	Q09644 caenorhabdi
11	199	3.1	1256	1 MRP_STRSU	P32653 streptococc
12	189	3.1	1902	1 P3P_LACLC	P15282 lactococcus
13	198.5	3.1	2249	1 OMPL_RICRI	P15921 rickettsia
14	197	3.1	1306	1 MSB2_YEAST	P32334 saccharomyc
15	192	3.0	1300	1 120K_RICRI	P14314 rickettsia
16	192	3.0	1654	1 OMPL_RICRI	Q53047 r outer mem
17	190	3.0	2710	1 TOXA_CLODI	P16154 clostridium
18	189.5	3.0	1645	1 OMPL_RICRI	P96989 r outer mem
19	189.5	3.0	2003	1 YDBA_ECOLI	P33666 escherichia
20	187.5	2.9	1288	1 VACA_HELPJ	Q9ZKw5 helicobacte
21	187	2.9	1902	1 P2P_LACLC	P15293 lactococcus
22	186	2.9	1902	1 P1P_LACLC	P16271 lactococcus
23	184	2.9	1661	1 APV_THETU	P38536 t amylopulli
24	182	2.8	1902	1 P2P_LACPA	Q02470 lactobacilli
25	179	2.8	1723	1 PM20_CHLPN	Q92812 chlamydia p
26	175.5	2.7	1858	1 P3K2_DICDI	P54674 dictyostell
27	174	2.7	1672	1 PMPB_CHLNU	Q09132 chlamydia m
28	173	2.7	2334	1 WAPA_BACSU	Q07833 bacillus su
29	172.5	2.7	1287	1 WACA_HELPJ	Q48245 helicobacte
30	172	2.7	1953	1 BIGA_SALT	P25927 salmonella
31	172	2.7	1802	1 HKRI_YEAST	P41809 saccharomyc
32	172	2.7	1849	1 IG4A_HAEIN	P45386 haemophilus
33	170.5	2.7	833	1 HSF_YEAST	P10961 saccharomyc

34	167.5	2.6	1076	1 NUPL_YEAST	P20676 saccharomyc
35	166	2.6	1256	1 ATL_STRAU	P52081 staphylococ
36	165.5	2.6	1291	1 VACA_HELPJ	Q48258 helicobacte
37	165.5	2.6	1778	1 N189_SCHPO	Q9UK4 schizosacch
38	165	2.6	1075	1 FLOS_YEAST	P38894 saccharomyc
39	165	2.6	1075	1 CBPA_CLOCL	P38058 clostridium
40	164.5	2.6	1464	1 NC02_HUMAN	Q15596 homo sapien
41	164.5	2.6	2358	1 YEEF_ECOLI	P76347 escherichia
42	164	2.6	1325	1 YDEK_ECOLI	P32051 escherichia
43	163	2.5	1228	1 SLAP_BACST	P35825 bacillus st
44	163	2.5	1419	1 ALAI_CANAL	Q13368 candida alb
45	163	2.5	1537	1 FLOI_YEAST	P32768 saccharomyc

ALIGNMENTS

RESULT 1
MGPC_MYCGE STANDARD: PRT: 1052 AA.
AC P22747: Q49257: Q49284: Q49385: Q49481: Q49482;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MgcP protein precursor.
GN MGPC OR MG192.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=90060815; PubMed=2583522;
RA Inamine J.M., Loechel S., Collier A.M., Barile M.F., Hu P.-C.;
RT "Nucleotide sequence of the MgpA (mgp) operon of Mycoplasma
RT genitalium and comparison to the Pl (mmp) operon of Mycoplasma
RT pneumoniae.";
RT Gene 82:259-267(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Peterson S.N., Smith H.O., Hutchison C.A., III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RT Science 270:397-403(1995).
RN [3]
RP SEQUENCE OF 60-133, 260-370, 441-512 AND 964-1052.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A., III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RT J. Bacteriol. 175:7918-7930(1993).
RN [4]
RP SEQUENCE OF 769-964 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=92051396; PubMed=1945886;
RA Peterson S.N., Schramm N., Hu P.-C., Bott K.F., Hutchison C.A., III;
RT "A random sequencing approach for placing markers on the physical map
RT of Mycoplasma genitalium.";
RT Nucleic Acids Res. 19:6027-6031(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>)

or send an email to license@slb.ch.

CC -----

DR EMBL: M31431; AAA25421.1; -

DR EMBL: U39698; AAC71411.1; -

DR EMBL: U02124; AAD12399.1; -

DR EMBL: U02161; AAD12443.1; -

DR EMBL: U34967; AAA88890.1; -

DR EMBL: U02157; AAD12439.1; -

DR EMBL: U34970; AAA88899.1; -

DR EMBL: X61525; CAB98130.1; -

DR EMBL: X61527; CAB98131.1; -

DR PIR: C64221; C64221. -

DR TIGR: M6192; -

DR Pfam: PF05220; MGPC; 1.

KW Cell adhesion; signal; Membrane; Complete proteome.

FT CHAIN 1 25 POTENTIAL.

FT CHAIN 26 1052 MGPC PROTEIN.

FT CHAIN 60 68 LDSSYQAD -> MIKPLPLLS (IN REF. 3).

FT CHAIN 125 126 TM -> SV (IN REF. 3).

FT CHAIN 474 481 TNNGTGP -> SNQISSGT (IN REF. 3).

FT CHAIN 476 476 AAD12439).

FT CHAIN 897 902 N -> S (IN REF. 3; AAA88890).

FT CHAIN 1052 AA; 114360 MW; F6C6D3B5D7789145 CRC64; SVSPRT -> MSHQGS (IN REF. 4).

SO SEQUENCE

Query Match 9.3%; Score 598.5; DB 1; Length 1052;

Best Local Similarity 23.7%; Pred. No. 9.8e-24;

Matches 270; Conservative 166; Mismatches 409; Indels 295; Gaps 49;

260 PIAOQKT---TTDNQNSTPFGSCAMPANRRYD-SOLNVRKIKTSFOLDERTINTNSN--- 313

21 PLALNTFLVEKDESKNTATYTPPTPTDTSK--SDLVSLAQDSSQIADQTHNTLNFV 78

314 RIGRRNNNSKFKVIGVDN-----PAHVIRETDGTFKFNFTN-OTQGEIYNDFL 361

79 LFESDVKVKESSGNNISFSDTSQGEKPSYVEFT-----NSTNIKIKTMKKYQL 132

362 DAPILPKDLHPDMYNYLTKRILPDVNTAVPWPGRVSGTNADGMPDCNGQITNTD 421

133 DVPNVSSDMQVNLKLEQPLTKYTLNSS-----LAKKQKQRE-VHLSGG----- 179

422 PIAOQKTTDNQNSTPFGSCAMPANRRYD-SOLNVRKIKTSFOLDERTINTNSN--- 474

180 ---QANQMTSQNRQNDLNNPSPANSSTGFKLTGNAYRKLSSEPTIEPTDGTQKQKMD 236

475 ---WTGSEEN--KNITRLATGSLPSN-----ERYWLDIPGTQVYTLKEDSV 516

237 SSGWSSTENKAKNDAPSVSGSSGTFNKYLTQALBSIGILFDQDPR----- 288

517 NVFSRLYLNSVNSLFIQDSIYIFGTSELPISLWYSEPTRLSDLTALNQKTDIEASST 576

289 NVITQLYASTSKLAVTNHILYVGNLFLPSMWYV-----VERSQAQ 330

577 DNGTGTGTTTADISSG-----STGAGTGTNTNT-----SQTVENPLNTYRS 620

331 EN--ASNKPTWFAFNTLMDGEDKOKOFEVNOIGYETITSTSHNFHSSTFOP---AYLI 385

621 FGID-----SKPTSAKIDETWADPNVTEARLYAEYRIGI 656

386 SGIDSVNDQITFSGFAAGSVGYDSSSSSSSTFQDQALASTTTSK--TCYKDLV 443

657 ONEIPTNAGNFIKRTIGVGYFTSGSRVYLASVNGDORPGNQPFLYVGYGYGQOT 716

444 TNDGLNGPIN-----GSFSIDTFESFVY--PYSGNHTNNGTGPITKATAYPVKKDKS 494

717 RTGFWYGYTK--LLNNSPYDVLDSRVGT-----ETNQFRRTSLTYPV-MGQY 762

495 -----YKINSINATPLNSTYGDGIEVPALGLANFNKSNQERLPSRTDQIFYGCI 546

763 LTBEGARSFNTPIRAQDTPESRSIFQSGYSNDTYEIQSVLGDGIRNNLANVYRS 822

547 VSPNELRS-AKSSADSTSGDTKVWMSNTQSRYLFPVPYNSGEGIIDADGKRPENRGASVT 605

QY 823 SFLNSRPNPNCLEMTAATTYLR-----QIGLARTSGL-PNQOPG--THQVIVSVSPGD 875

DB 606 TF-----SGLKSTAPDGFANSTANFSVGL--KAGIDPNVMSGKANKANGAVLTRGG 655

QY 876 QFSIKNIRTF--PGNOLWYFLFTNENN-----KS 904

DB 656 V-----VRLNPNNGND--SLSTTDNNIAPISFSTPTAAESAVDLTTEEVYNOES 707

QY 905 SVYTLRLADSNPDASSFSPT-----SLIDVNEIGVILLPLDINSFYTVNAGN-VALE 957

DB 708 GIMSEYFDSSLRKPSHDKQTPVDNMGFSYITYSRGIENL-QDAQFTLLDVAAPALAVQ 766

QY 958 SSNPGSPGTYAV-----NTEPNLSDI-----APE-----CS 985

DB 767 SGIGSTTQTLGLVPLSEFEFSAVIAKDSQNDKIDIKNNNGLEIDITQLSNSVATNGGL 826

QY 986 GAKYTS---DWGTFIOFPPDEYLLQNGFTSQVAREFTVNSPFLNSLVDTFPAAGTNYRV 1042

DB 827 APSYENRVDAMGKEFADNSVLAARNLVKTVDIELNTPELINSFREFPAPEDOKATL 886

QY 1043 VYDPPGNLTNQLPLKVOIOYLDGKYVDKLNKNNLVTFSYN-----NFGALPSWY 1093

DB 887 VATKQ---SDTSLVSPRIQFLDGNFYDL--NSTIAGVPLNIGFSPRVAFAGFALPAWY 940

QY 1094 VPPAIGSTIGILATMIIIGLAIQIPLRAQRKLDQKGFETTKKVDLTAAVGSYKKIIT 1153

DB 941 IPVSAGSSGILFILLVLGLGIGIPMYVRKLDQASFNVFKKVDLTAAVGSYKKIIT 1000

QY 1154 QFANVKKRPAALGAGSGDKRPAAPAKPAPKAPASSPAPPTGPKSGAPKPPAPK 1213

DB 1001 QTVGVKKAAPSALKAPNSVKKPAFLKP--PVQP-----PSKREGQKAVEYKSEETK 1051

RESULT 2

MGPC_MYCPN STANDARD; PRT; 1218 AA.

ID MGPC_MYCPN 050341;

AC 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 28-FEB-2003 (Rel. 41, Last annotation update)

DE MGPC protein precursor

GN MGPC OR MPN142 OR MP012.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_Taxid=2104;

RN [11]

RN SEQUENCE FROM N.A.

RP STRAIN-ATCC 29342 / M129;

RC MEDLINE-89211947; PubMed-2468577;

RX Inamine J.M., Loechel S., Hu P.C.;

RA "Analysis of the nucleotide sequence of the p1 operon of Mycoplasma pneumoniae."

RT RT Gene 73:175-183(1988).

RL [12]

RN SEQUENCE FROM N.A.

RP STRAIN-ATCC 29342 / M129;

RC MEDLINE-97103865; PubMed-8948633;

RX Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,

RA Hermann R.;

RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae."

RL Nucleic Acids Res. 24:4420-4449(1996).

RN [13]

RN IDENTIFICATION BY MASS SPECTROMETRY.

RP STRAIN-ATCC 29342 / M129;

RC MEDLINE-21088919; PubMed-11271496;

RX Regula J.T., Ueberle B., Boguth G., Goerg A., Schaefer M.,

RA Herrmann R., Frank R.;

RT "Towards a two-dimensional proteome map of Mycoplasma pneumoniae."

RL Electrophoresis 21:3765-3780(2000).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC
CC EMBL, M21519; AAA8336.1; -
CC EMBL, AE000002; AAB95660.1; -
CC PIR: J50069; J50069.
CC Pfam: PF05220; MGPC.1.
CC Cell adhesion: Signal: Complete proteome.
CC SIGNAL
CC FT CHAIN 1 25 POTENTIAL.
CC SEQUENCE 1218 AA; 130456 MW; 4DA29BCE41538311 CRC64;

Query Match 8.7%; Score 556.5; DB 1; Length 1218;
Best Local Similarity 22.1%; Pred. No. 1.8e-21;
Matches 321; Conservative 176; Mismatches 483; Indels 473; Gaps 65;

QY 1 MNISKLKSYTLIGLAVGALGSAPFGKQSDKNDNTOLVNOARTLDANSVRLAGLG 60
DB 1 MMSKRLKRYLLFLPLPLGLTSLA-----NTYLLQDHNTLTPY----- 40
QY 61 NGSLENTVLRVDYDNFTAAAGT-I-KLDSFTKPLYGLDSDCGGYKQIVSDYTSR 119
DB 41 -----PFTPLNGGLDVVRAHLHPSEYELDMKRVGDTKLVALVRS-ALVR 85
QY 120 NREDFORATYALLVNDDEANVHLKRINTNSNRIGNRNNSKFYIGVDNPAHVIRFTD 179
DB 86 VFODTTS-----SDQS-----NTNQLALSDPTQESOKALNG-----SOS 120
QY 180 GTFNFNTOTGEIYVNDFTLDAPILPKDLHPDMYNLYIQRLPLNDVNTAVVWPVGRS 239
DB 121 G-----SSPTSSNSODFASIVYLIFKAAPRATW--VEPKI-----KLALPY-VKQSS 165
QY 240 GTNADGMFDCNGOITNTDPIAQTQT--TDN-----QNESTENSGAMPGANR 287
DB 166 QSSGQGSNGKSLKTYTLDLLEQVPVPTPNAGLARVNGVADTVHFGSGESSMSNQ 225
QY 288 YSOLNVKH-----RIKTSQFLDERINTNSNRIGNRNNSKFYIGVDNPAHVIRFTD 343
DB 226 -RSQGLKNNPKPKAVTGFKLDK-----GRAVRLKNEWPVEYELD----- 266
QY 344 KENFTNOTGEIYVNDFTLDAPILPKDLHPDMYNLYIQRLPLNDVNTAVVWPVGRS 402
DB 267 -----TKEGKG-----KD-ESSMKN-----SEKTAENDAPLYGVMS 298
QY 403 TNADGMFDCNG-----QITNTDPIAQTQTNTDNPSTFN----- 439
DB 299 GAAGSASSLOGNSNGSLKSLRSAPVSVPPSTNSQTLISNPAPVGPQAVVSOQAG 358
QY 440 SGAMPGANRYSQNLVHKRIKTSFQDLKFF-----YPMWTSSEKNTTRLATGSLPN 495
DB 358 ARAAVSVNTASDTATFSKYTLNTAQLHOMGVIVPELKGWGN-----GTGVASR 410
QY 496 ERYWLDIP-----GTPQ-----VTL-----KEDSV-NVES 520
DB 411 QDATSNLPHAGASQGTGLGTSREPRALATSQRAVTVYAGPLRAGNSSEFDALPNVT 470
QY 521 RLYLNSVNSLSTIGDSIYFGTSELSLWY-----SEPTRLSDLT 561
DB 471 QLYHTSTQALAYLNGIYVWGS DRVSLWYVWVGDEQSGKATWAKTELMNGTDKOKOF 530
QY 562 ALNOKV-TDDIASSSTNDGTTNGTTTADTSSG-----STGA-GTGNTWT 606
DB 531 VENQIGFDDSDNSRLKAQGLQPAHYLLAGLDVADHLVFAAFKAAGVDMTTS 590
QY 607 SOTVSNPLNTYRSFGIDSKPTSAKIDETNNADPNVIEARLYAEYRLGIONEIPITNG 666
DB 591 SASSTYNQALAMSTTAGLSD-----GGYKALVENTAGLNGI 627
QY 667 N-----FIRNTIGGVFTSTGSRVVLRASTYNGDORPPNGFOPFLVFEYLGQOT 716

DB 628 NGLFTLLDTFAVVTVPVSGMKGSGQNNNEVQTTYPVSKDOKATAKIA----- 673
QY 717 RRGTFWYGYKTLKLNNSP-----YDV-----LDSPRGTETNORRSLNYP 757
DB 674 -----SLINASPLNSYDDGVYVDALGLNFKLNFKLEBRLLSRDQL---LVY- 718
QY 758 VMGGLYTE-----EGARSPSN-----TPYIRAGDPTES-----RS 788
DB 719 ---GIYNESELKSARRNAQSTSDDNSNTKVKYNTNASHLPYPPYYS-ANPEAGRRRA 774
QY 789 IFQSGYSDNTYE-----YIGSVLGFQDGRNNLNGVKASFSPLNSRPNGLIEMIAAT 842
DB 775 EDRNGVKIISTLESQATDGFANSLNCG---TGKAGVDPAAPARGHKPNYSALLVIRG-- 829
QY 843 YLRSGIARTSGILPNOQ-----PFGTTHQVIVSVSGDGFSSIKNR 884
DB 830 -----GVVRLNPNPDTKLLDSTDKNSPISFSTYPPFSASBAYDL-----TTLKDV- 876
QY 885 TIFPGNQLWYFLFTN-----ENKS-SVYTL-RLADSSNPASS---SFSEP 926
DB 877 TYIAESGLMFYTFDNGEKRTYDGKQGVKRRGAVIITSRGIENEDANTTLLSQAPA 936
QY 927 SLIDNEIGV-----ILPLDNEFTYVNA-----AGNALFSSNPGSGSTYAVTFNO 975
DB 937 ALAVQNGIASODDLTGILPLSDEFSAVITKDQTMGKYDIKN--TNGLEKDDQLSE 993
QY 976 N-----LSDAFESSGAKYTSDFWGTIOFRPDELIONGFTSOVARFVYNQSLNL 1028
DB 994 NVKRRDNGLVPIYNEG-----IYDIMGKRVDFANSVLQARNLTDKVDVEYINNPDLQSF 1048
QY 1029 VDETPANAGTNYRVVVDPPDGNLTNQLPLKVOIQYLDGKRY--DAKLK---NNNLVPS 1082
DB 1049 EKFTPAFDNORAMLVGEKTSDDT---LYAKPRIEYLDGNFYGDSKIAIPLINDPSRI 1105
QY 1083 YNNFGALPSWVVPATGSLGLIAMIILGLAIGLPLRAORLQDKGFTTEKKVDTLA 1142
DB 1106 FAGFALPSPVIVPVSQSSVGLIILLIIGLIGIPMYVVRKODSSFDVVKYDTLTP 1165
QY 1143 AVGSVYKRIITOTANYKKRPAALGAKSSDKRPAARPAAPKAPASSPAKRTGK 1202
DB 1166 AVGSYKRIITOTANYKKRPAALGAKSSDKRPAARPAAPKAPASSPAKRTGK 1210
QY 1203 SGAPTKPTAPKPA 1215
DB 1211 ---PVQP--PKRA 1218

RESULT 3
ADP1_MYCGA
ID ADP1_MYCGA STANDARD; PRT; 1122 AA.
AC 049379; 049437; 053351;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein).
GN MGCL.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=56;
RX MEDLINE=96201559; PubMed=8613358;
RA Keeler C.L. Jr., Hnatoow L.L., Whetzel P.L., Dohms J.E.;
RT Cloning and characterization of a putative cytodhesin gene (mgcl)
RL from Mycoplasma gallisepticum.
RN Infect. Immun. 64:1541-1547(1996).
RP SEQUENCE OF 1-12 FROM N.A.
RC STRAIN=56;
RA Hnatoow L.L., Keeler C.L. Jr., Tessmer L., Dohms J.E.;
RN Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.
RL [3]

RP SEQUENCE OF 378-570 FROM N.A.
 RC STRAIN-S6:
 RC MEDLINE-93371270; PubMed-8363503;
 RA Doms J.E., Hnawo L.L., Wenzel P., Morgan R., Keeler C.L. Jr.;
 RT "Identification of the putative cytohesin gene of Mycoplasma
 RT gallisepticum and its use as a DNA probe".
 RL Avian Dis. 37:380-388(1993).
 RN [4]
 RP SEQUENCE OF 159-1122 FROM N.A.
 RC STRAIN-S6:
 RA Goh M.S., Geary S.J.;
 RC Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
 RL -1- FUNCTION: COULD BE INVOLVED IN CYTADHERENCE.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
 CC -1- CAUTION: REF.4 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 159 TO 213.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U34842; AAB02987.1; -;
 DR EMBL; U44804; AAC83385.1; ALT_FRAME.
 DR PIR; J118346; T18346.
 KM Cytoadherence; Signal; Transmembrane.
 FT SIGNAL 30
 FT CHAIN 31 1122
 FT TRANSMEM 997 1021
 FT DOMAIN 1002 1009
 FT CONFLICT 313 314 DM -> IW (IN REF. 4).
 FT CONFLICT 382 382 Y -> F (IN REF. 3).
 FT CONFLICT 501 501 I -> T (IN REF. 3).
 FT CONFLICT 568 568 R -> G (IN REF. 3).
 FT CONFLICT 570 570 T -> A (IN REF. 3).
 FT CONFLICT 693 695 DIL -> VIT (IN REF. 4).
 SQ SEQUENCE 1122 AA; 121305 MW; 155C34DA2D6C3C65 CRC64;
 Query Match 7.8%; Score 501.5; DB 1; Length 1122;
 Best Local Similarity 22.6%; Pred. No. 1.2e-18;
 Matches 311; Conservative 174; Mismatches 470; Indels 423; Gaps 68;
 OY 5 KRKISYTLIGLAVFAGLSASFQKSDKNDMTQVYNQARTLDANSVRLAGLGNGSL 64
 DB 2 KKL-TFKLSVGTPLALIGLSFGLAVSGARPNMLKPVNOVGEMNSQ-----GQS-NL 52
 OY 65 FNTVLADVDFN-----ITFANGTIIKLDSFTKPLVG-----LDLS 100
 DB 53 LEKARWRNSNFTSLSIDGTNPGALVILGSKSISRID-----LVGNVIMTFPGNTNDLT 107
 OY 101 DDCGGY-----KVKQIVSDY---TTSRNFQROIRAYVALLVNDENVHLKRI 146
 DB 108 GKGVEFDANNKLLAFSGDVSFNVSDLSKTVVEATODEBDENVFILLMPAA----- 160
 OY 147 NTSNRTGNNNNSKRVIGVDNPAHVIRFTDDGTRKFNFTNGEYI-NDFFILDAPLLP 205
 DB 161 -----VQEQKTKQVENEYFMSDAP--- 181
 OY 206 KDLHPDWNLYIORKLLPNDVNTAVVPWVYKVSIGNADDG-----MFDGDN 252
 DB 182 -----AAGDTSAGSATPAGGSGSAGGAVAPAAASSTARLVEGDN 225
 OY 253 -----GQTTNDPLAQT-----KTTTNDQNSTFNSGAMPGANNRYDSQLNV 294
 DB 226 SAGNGTMTPTASTSEIYIDVNSDNKIPKTKTLDSESESEINGRTYAN--INTOKUL 283
 OY 295 KHRKTSFOLDERINTNSNRIGNN--NNSKFV--IGVDN-PAHVIRFTDDGTRFN- 346
 DB 284 GGVV-----VKVENLNFENSENPFVAVENMAFIRPKDQVNDVPTWTGGSANGKMTNV 334

OY 347 --FTNTOGEIVNDFILDAPIILPKD-----HPOWNLXYIQRIKLPND 387
 DB 335 LQFYKHDPNPAVNNRFYRAKYRKLETOITPPLIDSSFPYEHPEMW-----E 383
 OY 388 VNTAVVMP--VGRVSGTNADGMDFC--GNGOITND--PLAQTKTTDNO-----N 434
 DB 384 GNOFVMPWMQYITNLGLYAKDGWVYLFGGNGWVNNESALSIGVFRTEFKRTAEACN 443
 OY 435 PST--FNSGAMPGANNRYDSQNLNKHRIKTSFQDKEKFEYEWGSEENKNTIRLATGSL 492
 DB 444 TKVGYPPGILLSAIS--FDATRNGLAIAALQDVCYHR-----VRLAVGCV 491
 OY 493 PSNERXYIIPGTPOVTLKEDSVNVSRLYNSVNSLSITGSIYIFGNS--LPSLM 549
 DB 492 SSP-----RCANINIFLGSAI--TWGINGNFIPTDKW 521
 OY 550 YSFPTRLSDLTALNOVKTDDLEAST--DNGTTNGTNTTADTSSGS-----TGAG 599
 DB 522 H--SNAVIED-----APTFITYNSSGVLNNGSGQSTSPMPNSNGNESIPRYMTNSY 573
 OY 600 TGNNTTSQTVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVI--EARIYAEV----- 652
 DB 574 DYSNVPFALLISKPAAGWTKQ--VESLPTTAKLIDTLN-SLKPFTQENNIFSYALMD 630
 OY 653 ---RLGTONE--IPTNAGNFIRMTIGGCVFTSGSHVYLASYNG-----DORPTG 699
 DB 631 RQMSLQTRSDTWLTNT 690
 OY 700 NFQPFY-----VFGYLGQOQRTGTFW---YGYKLLNSPYVLDSPRGTEFNOFR- 750
 DB 691 NVD-ILYSNNTNFTYY--YQVGGAITTWPEVOVNKTSANITYNLTRTDGSTITPATQD 748
 OY 751 RSLTITPVMGYLTGEGARS--FSNTPYRAGQDPEPSRIFQSGYSDNTYEYIQSVLGF 808
 DB 749 ANTVYSKLNAYLSTGDDQGWYNGSIYKAKKSFPPSQ-----GYTMQDEKGLTTTASN 803
 OY 809 DGIRNNLVGVKASFLNSRPNPGLMIATTYIRSGIAGLARISGLPNQOPFGTTHQV 868
 DB 804 AVISNMTRKACY-----SIRPDD-----TFVSYSKI-----PF----- 831
 OY 869 ISVSPGQFQSIKNIIRTIIFPQNLWFLFTFNENKNSVYTLRLDSSNDPDASSFSPSL 928
 DB 832 -----EKETIAAVNRSLDS-----YYQLNGERTSVMTVAHV---SPDSSA----- 869
 OY 929 IDVNEIGVILPLDINSFYVNAAGNALFSSNPGSGSTIAVNTFNONISDLAFEGSGAK 988
 DB 870 LFLNPKRTITNLMNRD-----NVIGOGAFISRN-DIPSSF-----FENKINDIV----- 912
 OY 989 YTSDFWGTIOKRPDEYLIQNGFTSQVARNPVTNOSFINSLVDTFPA--NAGTYRVVVDP 1046
 DB 913 -----TTEDGTEYLDKTYINSIXRTTPQNNPHIRLRLVLVD 950
 OY 1047 DGNLTNOMLPKLVQIOYLDGRKYDYDAKLNNNLVY--FSYNNFGALPSWVYFATAGSLTIL 1105
 DB 951 RSRATNDFIKLLPQV-LVDEGVAVAPQANSYFVSDOETGTDALGVGLVALISIPILII 1009
 OY 1106 AITMIIIGLAIPLPAOKRLDQKGTKEFKFVDLTLAAGSVYKKTITQTA--NVKKRPA 1163
 DB 1010 ALALALGIGIGIPMSQNRKMLKOGFAISNKVDILTTAVGSEVFOIINRISVFIKKTPQ 1069
 OY 1164 ALGAGKSGDKRPAAPAAKPAAPKPSAPKASSPAKFTGSCAPKPTAP--KRAPKPT 1220
 DB 1070 MLOA-----KKKOGA-----SSPSKPSAPAKKFTGPTKP--SAPGAKPTAPAKKAPAPT 1118
 RESULT 4
 ADPL_MYCPN STANDARD; PRT; 1627 AA.
 ID ADPL_MYCPN
 AC P11311; 1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 17, Last sequence update)
 DT 01-FEB-1991 (Rel. 41, Last annotation update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Adhesin_P1 precursor (Cytadhesin P1) (Attachment protein).
GN MGPB OR MPN141 OR MP013.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID:2104;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN-ATCC 29342 / M129.
RX MEDLINE=88057593; PubMed=3119495;
RA Su C.-J., Tyson V.V., Baseman J.B.;
RT "Cloning and sequence analysis of cytodhesin P1 gene from Mycoplasma pneumoniae";
RT Infect. Immun. 55:3023-3029(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=88297153; PubMed=2841195;
RX Imamine J.M., Denny T.P., Loechel S., Schaper U., Huang C.H.,
RA Bolt K.F., Hu P.C.;
RT "Nucleotide sequence of the P1 attachment-protein gene of Mycoplasma pneumoniae";
RT Gene 64:217-229(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himelereich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RT Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";
RT Nucleic Acids Res. 24:4420-4449(1996).
RN [4]
RP SEQUENCE OF 1301-1520 FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE=88154763; PubMed=2450165;
RA Dello S.F., Su C.-J., Horton J.R., Baseman J.B.;
RT "Identification of P1 gene domain containing epitope(s) mediating Mycoplasma pneumoniae cytoadherence.";
RT J. Exp. Med. 167:718-723(1988).
RN [5]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE=21088919; PubMed=11271496;
RA Regula J.T., Ueberle B., Boguth G., Goerg A., Schnoelzer M.,
RT Herrmann R., Frank R.;
RT "Towards a two-dimensional proteome map of Mycoplasma pneumoniae";
RL Electrophoresis 21:3765-3780(2000).
CC -1- FUNCTION: THE PROTEIN IS THE MAJOR ADHESIN MEDIATING THE ATTACHMENT OF THIS MYCOPLASMA TO RESPIRATORY EPITHELIUM.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M18639; AAA25424.1; -;
DR EMBL; M21519; AAA88325.1; -;
DR EMBL; AE000002; AAB95661.1; -;
DR EMBL; X07191; CAB37298.1; -;
DR PIR; A41480; A41480.
DR PIR; S03725; IUYMAP.
DR InterPro; IPR004940; Adhesin_P1.
DR Pfam; PF03257; Adhesin_P1; 1.
KW Cyadherence; Signal; Transmembrane; Complete proteome.
FT SIGNAL 1 59
FT CHAIN 1 60 1627 ADHESIN P1.
FT TRANSMEM 1527 1547 POTENTIAL.
FT SITE 1403 1415
SEQUENCE 1627 AA; 176220 MW; 14F7A2CBA36E6116 CRC64;

[illegible]

```

OY 1016 RNFPT-----NOSFLNSLVD-ETPAN-----AGTIVRY-----VYDP 1046
DB 1404 POSVKEKSPDOIDENRLEFTHRYVDLFPDPTMLVYDQYIPLEFIDIPASVMPKMKVRLKLSF 1463
OY 1047 DGNLTNOMLPLKVOIQYIDKTYDAKLKN-----NLYVTS-----YANFGA 1088
DB 1464 DFN--EQLSLGLEFEKFPD---QDTPNNVQVNPNGDPLPLITASSQGPOTLFSPFNO 1518
OY 1089 LPSWVPVPAIGSTIGLIMILILGALIGPLRAQRKLDQKGFTEKKVDLTFAVGSY 1148
DB 1519 WPDYVLPATYIPYIVYLVSTLGLAIGIPMKHKQALKGAFALSNQKVDYLVKAVGSVF 1578
OY 1149 KRIITOTANVKKRPPALGAGSGDKKPAAPAKKPPAKSPKSPKPGKSGAPTK 1208
DB 1579 KEIINTT-GISQAPKRL-----KQTSAAKPGAPPPVPPKPGAKPPVQP-----PKK 1625
OY 1209 P 1209
DB 1626 P 1626

RESULT 5
ADP1_MYCSE STANDARD: PRT: 1444 AA.
ID ADP1_MYCSE
AC P20796; Q49286;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Adhesin P1 precursor (Cytadhesin P1) (Attachment protein) (MgPa).
GN MGPA OR MGPA OR MG191.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=89173298; PubMed=2925238;
RA Dailo S.F., Chavoya A., Su C.-J., Baseman J.B.;
RT "DNA and protein sequence homologies between the adhesins of
RL Mycoplasma genitalium and Mycoplasma pneumoniae";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=90060815; PubMed=2583522;
RA Imanue J.M., Loechel S., Collier A.M., Barile M.F., Hu P.-C.;
RT "Nucleotide sequence of the Mgpa (mgp) operon of Mycoplasma
RL genitalium and comparison to the p1 (mpg) operon of Mycoplasma
RN pneumoniae";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569933;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RL Fritschman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fleischmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RL "The minimal gene complement of Mycoplasma genitalium";
RN Science 270:397-403(1995).
RN [4]
RP SEQUENCE OF 106-177 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=92051396; PubMed=1945886;
RA Peterson S.N., Schramm N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RL "A random sequencing approach for placing markers on the physical map
RN of Mycoplasma genitalium";
RN Nucleic Acids Res. 19:6027-6031(1991).
RN [5]
RP SEQUENCE OF 468-588; 627-723; 980-1162 AND 1235-1289 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;

```

```

RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RL "A survey of the Mycoplasma genitalium genome by using random
RT sequencing";
RN J. Bacteriol. 175:7918-7930(1993).
CC -1- FUNCTION: THE PROTEIN IS THE MAJOR ADHESIN MEDIATING THE
CC ATTACHMENT OF THIS MYCOPLASMA TO THE CLILATED EPITHELIUM.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE ADHESIN P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M31431; AAA25420.1; -
CC EMBL: U39698; AAC71410.1; -
CC EMBL: X61522; CAA43734.1; -
CC EMBL: U02209; AAD12501.1; -
CC EMBL: U01694; AAB01007.1; -
CC EMBL: U01779; AAD10599.1; -
CC EMBL: U02159; AAD12441.1; -
CC PIR: A30588; A30588.
CC TIGR: MG191; -
CC InterPro: IPR004940; Adhesin_P1.
CC Pfam: PF03257; Adhesin_P1; 1.
CC Cytadherence; Signal; Transmembrane; Complete proteome.
CC KW CHAIN 1 1444
CC SIGNAL 31 1444
CC TRANSMEM POTENTIAL.
CC CONFLICT 1160 1161 SO -> FA (IN REF. 4).
CC FT SEQUENCE 1444 AA; 159651 MW; DBIDFF7A90FFFB8A CRC64;
SQ

Query Match 4.4%; Score 285; DB 1; Length 1444;
Best local similarity 20.9%; Pred. No. 2.8e-07;
Matches 270; Conservative 159; Mismatches 477; Indels 386; Gaps 68;

OY 24 SASGFEKQSD-----KSDNTQVQNAQRTLDANSVYLAGISGSLFNTVLRDQVDN 75
DB 418 NTSGFVFGDTHDKKDFKKNSSPTALPPEAFANIGMVAIG-----NSV----- 464
OY 76 FITANGTIIKLDSEFTKPLYLGLSDDCGKYKQIYSDYTSRNFQDQRTAV-YAL- 133
DB 465 FIFGNGHATKMT-TNPL-----SIGVFRK-----YT---DNFSKSYTGMPYAL 508
OY 134 ---LVNDEANHLKRLNTNSNR-----IGNRNNSKFVIGG---VDNPAH 172
DB 509 FEGLINPQTN-GLADPLGTRKFEYVPRMAVSGVKWGN-----OLVLAGTLTGMDTAT 562
OY 173 VIRETDDGSK--FNFNTQGEIVNDFLDAPILPKLHPDWNYIQRKILPNDVNTAV 230
DB 563 VRLKXYDQLEKHLNVAGOGGLREDLQIFTPY-----GMAV-----RDIP----- 605
OY 231 VPMYVGRVSGTADDCMGDCGQGTNTDTPAQTKTNDONPSTFGSGAMPAN---NR 287
DB 606 GAW-----LQEDM-----GSKFGPHYFLNN 625
OY 288 YDSQLNVKHKRITSFQDERINTNSRIGNRNNNSKPYIGVDPAHVIRETDDG---T 343
DB 626 PDIQDQVNN-----DTVALISSYKNDKL-----KHVYPRYSGLAWQ 665
OY 344 KRFNFTNQ-TOGEIVNDFILDAPILPKLHPDWNYIQRKILPNDVNTAVPMYVGRVSG 402
DB 666 LFNMSNKLNTPTLSANFVNDNSYAPRSLF-----AAILNEDLTLGL----- 706
OY 403 TNADDCMGDCGQGTNTDTPAQTKTNDONPSTFGSGAMPANRYSQNLVYKRIKT 462
DB 707 ---SDKIFYGKENEFEADRFNQLSLNPNP-----NTNARYLINVQRTFT 752
OY 463 SFQD-----EKFV-YPMWIGSEENKNITRLALGSLPSNERYWILDIPGTPVYTLKESVN 517

```



```

Db      753 GPNLDSRFDOFLDFLPWIGN--GKPFNSPSPSTSSASS-----STPLPFPESNIING 802
Qy      518 VESRL--YNSVNSLSFIDSDIYIGTSLPSLM-----YSPPTRLSDLTALNOXYTD- 570
Db      803 VASMTQHLKNTNR-----WFLPNPSPDIMGAGRYVOSANOKNIPPEQVPSN 855
Qy      571 ---IEASSTDNGTNGTTTADT-----SSGSTAGTGNNTNTSOTVSNPLTN--TVR 619
Db      856 STPFDPNSDNKVTBPGSSSKPTTPALPNSISPTSDMNLALFTTK--NNPQRNOLLR 913
Qy      620 SFGIDSKPTSAKKIDFTN-----WADPNVIEARI--YAEYRLGIQNETIPITMAGN 667
Db      914 SL-LGTIPVLYLNKSGSDNQFNKDSQKWDKTETNCGNLPGEVFN-GLYNALLHTYGF 971
Qy      668 FARNITIG---GVGF-----TSTGSRVLYRASYNQDQRPKN-----FQPELYVFGY 710
Db      972 FGTNTNSTDPKIGFADSSSSSTLVGGLMWTSDQVNLVLYNDTSFGFOLGGMFTTF 1031
Qy      711 LGYQOQRTGTFWVG--TYKLLNNSPYVDLSPRAGTETNQFRRTSLTYPMGGLYTFEGA- 768
Db      1032 TDFIRPRTG--YLGITLSSLOPOTIMADP-----WTSFK-----GSTLSDGTP 1075
Qy      769 -----RSFSNTPYIRAGDTPESRSIFQSGYSDNTYEXIQSVLGFQDGIKNNLNVGKAS 822
Db      1076 KSLMDPTALKSLPNSSTYDTNPTLSPFQLYQNKVAYQTNTY----NKLIEPVDAF 1131
Qy      823 SFLNSRPNPNLEMIATTYLRSQIGLARTSGLPNOQPFQTHQVIVSPGDQESIKN 882
Db      1132 SAAT---MTSLKLKLTTKTKAKLGKTAASQGNNGGVSQITNTTITGNISEGK 1187
Qy      883 ITIIPGNOIWFLETFNENKSSVYTLRLADSSNPASSFSFTSLIDNEIGVILPLD 942
Db      1188 EERSIGAETLKKFEPSKONKS-----EIGI-----GD 1215
Qy      943 NSFTYNAAGNVALESNGSPGSTYAVTFNQNLSIDIAEESGATYSDFMCTOPKPD 1002
Db      1216 SFTFKMD--GKLGAVSTP-----LVMLIN-----GGATSDSD--TEKISFKG 1256
Qy      1003 EYLIONG--FTSGVARNFTVNSOFLNS-----LYDFTPA--NAGTNRYVAVDPGNIJN 1054
Db      1257 NQIDFNRLFTLPTVELFDPRNMFVYDQVPLVNLPSGPDQASIRLKV-----STSVEN 1311
Qy      1055 LPLKQIOYLDGK-----YDPAKILNNMLVTFSYNNGFALPSWVPVTAIGSLTILAI 1108
Db      1312 QFLGVRLFEKDPQTQGFIPVLNASTSGPQTFEQPFNQMA---DYVLPILYTVIVYIIS 1368
Qy      1109 IILGLAIGIPLRQAKLQDKGRTTKKYVDTLTAANGSYKKITITOT--ANVKKKPAALG 1166
Db      1369 VTLGLTIGIPMHRNKKALQAGDLSNKKYDVLTAKAVSFEKELIINTGISNAFK----- 1423
Qy      1167 AGKSGDKRPAAPAAKPAAPKPSAKSPAKP 1198
Db      1424 -----LKQATPTKPT-PK--TPPKP 1440

```

```

RESULT 6
OMPA.RICCN          STANDARD:          PRT: 2021 AA
AC 052657; P95591; P95592; P95593; P95594; 052667; 052668; 052669;
AC 052670; 052674;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Outer membrane protein A precursor (190 kDa antigen) (cell surface
  antigen) (TompA) (Tomp A).
GN OMPA OR RC1273.
OS Rickettsia conorii.
OC Bacteria: Proteobacteria: Alphaproteobacteria: Rickettsiales;
  Rickettsiaceae: Rickettsia: Rickettsia.
OX NCBI_TaxID=781;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=Malish 7;
RX MEDLINE=94171067; PubMed=8125327;
RA Crocquet-Valdes P.A., Weiss K., Walker D.H.;
RT "Sequence analysis of the 190-kDa antigen-encoding gene of Rickettsia
  conorii (Malish 7 strain).";
RL Gene 140:115-119(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V.,
  Raoult D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
  RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
  Science 293:2093-2098(2001).
RN [3]
RP SEQUENCE OF 8-204 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RX MEDLINE=97015921; PubMed=8862558;
RA Roux V., Fournier P.E., Raoult D.;
RT "Differentiation of spotted fever group rickettsiae by sequencing and
  analysis of restriction fragment length polymorphism of PCR-amplified
  DNA of the gene encoding the protein rOmpA.";
  J. Clin. Microbiol. 34:2058-2065(1996).
RN [4]
RP SEQUENCE OF 953-2012 FROM N.A.
RC STRAIN=Indian tick typhus, M1, Malish 7, and Moroccan;
RA Raoult D., Fournier P.E., Roux V.;
RT Phylogenetic analysis of spotted fever group rickettsiae by study
  of the outer surface protein rOmpA.";
  Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
RL -1- FUNCTION: ELICITS PROTECTIVE IMMUNITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
  S-LAYER WITH HEXAGONAL SYMMETRY.
CC -1- PTM: GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 001028; AAA17405.1; -
DR EMBL: AE008674; AAL03811.1; -
DR EMBL: 043794; AAB49548.1; -
DR EMBL: 043798; AAB49550.1; -
DR EMBL: 043806; AAB49551.1; -
DR EMBL: 045244; AAB49566.1; -
DR EMBL: 046918; AAB66663.1; -
DR EMBL: 083440; AAC35176.1; -
DR EMBL: 083443; AAC35179.1; -
DR EMBL: 083448; AAC35184.1; -
DR EMBL: 083453; AAC35189.1; -
DR InterPro: IPR006315; Autotransport.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRfams: TIGR01414; autotrans_bail; 1.
KW Antigen; Repeat; Signal; Cell wall; S-layer; Glycoprotein;
  Complete proteome.
KW -----
FT SIGNAL 1 38
FT CHAIN 39 2021
FT DOMAIN 238 946
FT DOMAIN 1424 1528
FT VARIANT 60 60
FT VARIANT 76 76
FT VARIANT 86 137
FT VARIANT 126 133
FT VARIANT 953 954
FT VARIANT 1245 1245
FT VARIANT 1308 1308
FT VARIANT 1308 1308

```

```

POTENTIAL.
OUTER MEMBRANE PROTEIN A.
THR-RICH.
THR-RICH.
N -> N (IN STRAIN INDIAN TICK TYPHUS).
R -> H (IN STRAIN INDIAN TICK TYPHUS).
MISSING (IN STRAIN M1).
MISSING (IN STRAIN MORCCAN).
VT -> II (IN STRAIN INDIAN TICK TYPHUS).
D -> A (IN STRAINS INDIAN TICK TYPHUS), M1
  AND MORCCAN).
N -> H (IN STRAIN MORCCAN).

```

Query Match 3.68; Score 232.5; DB 1; Length 2021;
 Best Local Similarity 21.0%; Pred. No. 0.00023;
 Matches 278; Conservative 147; Mismatches 536; Indels 365; Gaps 64;

```

  QY 9 SYTLGGTAVFGALGSASFQSGKSDKSDNTOLYONAKRTIDANSVRLAGLGONGSLFNTV 68
  DB 162 NTYTGALNALGSA-----NAALLIGSAA--PAKITLAG--NINGGGITIV 202
  QY 69 LRDVDNFTTANGTITLDSFTKPLYGLDSDCGYKVKOIVSDYTSRNFQDQCR 128
  DB 203 KTD-----AATNGTIGTNTALATVYVAGIA--FLEGAIKATTTKLTNNAVLTITNV 255
  QY 129 AYAALVNDNAV--HLKRINTN-----SNRIGRN-----NSKFEVIGVDNPAHVI 174
  DB 256 AVLTGALDNTGVDNGLNLGALSQVYTGNTGNALATISVAGAKATLGAVIKATTT 315
  QY 175 RFTDDGKFNFTNQ--TQGEIYNDFLDAPILPKDLHPDWXNLYIQRKILPNDVNTAVVP 232
  DB 316 KLTDNASAVFTNPVVVTGALDN-----TGANNNGIVT 348
  QY 233 WP-----VGRVSGTNADGMEFGDGNQITNTDPIAQTCT--TDNONPSTF--NSGAMGA 284
  DB 349 FTGDSYVTGNTGNTNA--LATISVAGAKATLGAIKATTTKLTNNAVFTTNPVVVTGA 407
  QY 285 NNRYSQNLVHKRIKTSFQIDERINTNSNRIGRN-----NSKFEVIGVDNPAHVI 337
  DB 408 ---IDNTGNANNNGIVT--FTGDSYVTGN--IGNTNTALATISVAGAKATLGAIKATTTK 460
  QY 338 FTDDGTFENFTNQ--TQGEIYNDFLDAPILPKDLHPDWXNLYIQRKILPNDVNTAVVP 395
  DB 461 LTNASAVFTNPVVVTGALDN-----TGANNNGIVT 493
  QY 396 P-----VGRVSGTNADGMEFGDGNQITNTDPIAQTCT--TDNONPSTF--NSGAMGA 447
  DB 494 TGDSTVYTGNTGNTNA--LATISVAGAKATLGAIKATTTKLTNNAVFTTNPVVVTGA 551
  QY 448 NRYDSQNLVHKRIKTSFQIDERINTNSNRIGRN-----NSKFEVIGVDNPAHVI 507
  DB 552 --IDNTGNANNNGIVT--FTGDSYVTGN--IGNTNTALATISVAGAKATLGAIKATTTK 589

```

QY 508 QVTLKEDSVNVFSRLYNSLSFTGDSIYIFGHSLEPLMYXSFPTRLSDLTALNOVK 567
 DB 590 KATLGAIKATTTKLTNNAVFTTNPVVVTGALDN-----TGNTNNSQVNSPTL 615
 QY 568 TDDIEASSTNDNGTGTGTTTADTSSGTYAG-----TGNTNNSQVNSPTL 615
 DB 630 ANNGIVFTTNNSVTGNTGNTNA--LATISVAGAKATLGAIKATTTKLTNNAVLTITNV 689
 QY 616 NTVRSFGIDSKPTSAKIDETNMADPNVIEARIVAEYRIGIONEIPITNAGFNRTTGG 675
 DB 690 NAVLTGALDN-----NTGVDNGLVNLGALSQVYTGNTGNALATISVAG--KATLG 742
 QY 676 VGFSTGSRVYLRAS-----YNGDQRPNGF--QPELY 706
 DB 743 AVIKATTTKLTNNAVFTTNPVVVTGALDNNTGNANNNGIATFGDSYVTGNTGNALAT 802
 QY 707 VEGYLGYOQRTGTFWYGTIKLNSPYDVLDSPRV-----GTE 745
 DB 803 VNVGAGLLRVGGVAKSNTINLTNNAVFTTNPVVVTGALDNNTGNANNNGIATFGDSY 862
 QY 746 TNGFRRTS--LTYPMGGYLTEGARSFNTPYIRAQDTPP--SRIFQSGYSDNTYEYI 802
 DB 863 TGNIGTNTALATISVAGAKATLGAIKATTTKLTNNAVFTTNPVVVTGALDN----- 918
 QY 803 QSVLFGDGRNNLNVGVASSPLNSNRPNGLMIATYLRISQIGLARTSGL--PNOQP 861
 DB 919 -----GNANNNGIVFTGDSYVTGNTGNAL--ATVNVGAGVTLQAGGSDANND 967
 QY 862 FGTTHOVIVSPQDQFSSIKNRTIFPGNOLMYFL--FTENENK-----SVYTLR 910
 DB 968 FGARSFLFENGPLDG-----GGNALPYEYKGAIVANGNAILNVTKLLTAVHLT 1016
 QY 911 LADSS--NPDASSF-----SPTSLIDVNEIGVILPLDNSFEYVNAAG--NVALSSNP 961
 DB 1017 IGVVAELINGAGLFLIDNASAGDVITLNMODIHFAFALSAVLSTLGVGNVNTLLADL 1076
 QY 962 GSPGSYAVNTFENQNL-----SDIAFESSGAKITSDFWGTIOFQEPDEYILONGFTSOVA 1015
 DB 1077 VAPGVDEGIVVFDGAVGNLIGISNVA--GAARNIGDGN--KFNILLIYNVATTIDD 1130
 QY 1016 RNEVTNOSFL--NSLVDPTPA--NAGT-----NRRVVDPDGNLTNOMPLKVOIQVLD 1065
 DB 1131 VNLGIGNVLLNNADDTSTTAFNAGTIQINDATYTI--DANNGNLIMPAG--NIKF-- 1183
 QY 1066 GKYYDAKL-----KNNNLVTFSTYNNFGALPSVNVFTAGISLTGLAIIMI--LGLA 1114
 DB 1184 -AHADAQLIILDNSSGNDRTITLGN-----IDPD--NDDEGIIVILNSVYAGKKLITA 1232
 QY 1115 IGIPLRARKLQDGGFK-----TFPKVDTLFAVGS-----VYKKII-- 1152
 DB 1233 GKRTFGAKHLODIVFKGSDGFTAGTTFNTIVLIDITGOLBGLATFANVLFKDAVOL 1292
 QY 1153 TOTANV 1158
 DB 1293 TOTGNI 1298

RESULT 7
 OMPB_RICPR STANDARD: PRT: 1643 AA.
 ID OMPB_RICPR
 AC 053020; 09ZCM0;
 DT 30-MAY-2000 (Rel. 39, last sequence update)
 DT 30-MAY-2000 (Rel. 41, last annotation update)
 DT 28-FEB-2003 (Rel. 41, last sequence update)
 DE (Surface protein antigen B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen B precursor (168 kDa surface-layer protein)
 DE (antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 DE (antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
 GN OMPB OR SPA OR SP4 OR RP704.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OC NCBL_TaxID=782;

Db 890 -----IKSTVG-NGVDNENAPLIVVSGIDSM----- 915
 QY 834 GLEMIAATYLRSGIGARTSGLPNOOPGTHQVIVSPGQOFSSIKIRIFEPGNOLW 893
 Db 916 -----INNGQIIIGDKKRIIALISGDSNSTTANANTLYSG----- 949
 QY 894 YLEFTNENKSSVYTLKADSSNDP-----ASSSFSP-----TSLIDVEIGVILP--LL 941
 Db 950 --IRFTKNNQGTV--TLSGGFENPNTGTIYGLGLENKSPKLVQVFTDYNNLSIIANNVT 1006
 QY 942 DNSFTY-----VNAAGNALFSSNPGSPSYTAVNTFNONLSIDAF 982
 Db 1007 INDYTLTGTGGIAGTDFDAKITLGSVNGNANVRFVSTFSDPSMVAIQANK----- 1059
 QY 983 EESGAKYISDFWCTIOEFKPEDELLQNGFTSQVARNFYNOSFLN---SLVDFTPAAGTN 1039
 Db 1060 -----GIVTY-----LGNALVSNIGSLDTPVASVRFGTGDSGAG 1093
 QY 1040 YRVVDDPGLNLTNOLPLKVOIQYLDGKYDAKLNKNLNV---TFSYNN----- 1085
 Db 1094 LQ-----GNIVSON-----IDFGTYNLTJNSVITLGGTTAINGEIDLLTNLI 1138
 QY 1086 -FGALPSPVPAIGSTLGI 1104
 Db 1139 FANGSTWGDNISITTLNV 1158
 RESULT 8
 OMPB_RICJA STANDARD: PRT; 1656 AA.
 ID OMPB_RICJA
 AC 00653:2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Outer membrane protein B precursor (168 kDa surface-layer protein)
 DE (Surface protein antigen) (Cell surface antigen 5) (Scs5) (rompB)
 DE (romp B) (contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide).
 GN OMPB
 OS Rickettsia japonica.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiase; Rickettsia.
 OC NCBI_Taxid=35790;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-YH:
 RA Uchiyama T.;
 RT "Sequencing of the gene encoding the protein romp B of Rickettsia japonica";
 RL Submitted (MAY-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY SIMILARITY).
 CC -1- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-LAYER WITH HEXAGONAL SYMMETRY.
 CC -1- SIMILARITY: BELONGS TO THE RICKETTSIAL OMPA/OMPB FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions in no way use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AB003681; BAA20138.1;
 DR InterPro: IPR006315; Autotransport.
 DR InterPro: IPR005546; Autotransporter.
 DR Pfam: PF03797; Autotransporter; 1.
 DR TIGRPFAM: TIGR01414; autotrans_bar1; 2.

KW Antigen: S-layer, Cell wall. 120 kDa SURFACE-EXPOSED PROTEIN.
 FT CHAIN 1 1338 32 kDa BETA PEPTIDE.
 FT CHAIN 1339 1656 POLY-GLY.
 FT DOMAIN 528 533
 FT SEQUENCE 1656 AA: 168097 MW: 3132A69C9D5999F CRC64;
 SO
 Query Match 3.4%; Score 215; DB 1; Length 1656;
 Best Local Similarity 19.0%; Pred. No. 0.0014;
 Matches 285; Conservative 161; Mismatches 492; Indels 562; Gaps 74;
 14 GGLAV-----FGALGSASFGRKOS---DKSNDNTQLVNQAARTLANSVRLAGLQNGSLF 65
 Db 160 GGAAIANDLSDGIGTIDFGAASSTLVDFLANPTQ---KAPILADNALIYN-GANGTL- 214
 QY 66 NTVLRVDNDFITRANGTIIKLDSFTKPLPYGLDLPDCCGKYVKQIVSDYTTSRNRPEQR 125
 Db 215 -----NVTNCFIOVSDKSPATYKAIIGDQO----- 241
 QY 126 QTRAYVALLVNDEANVHLKRIINTNSRIGRRNNSFFVIGVDNPAHVIRETDDG--TKF 183
 Db 242 -----FMFTNA-----TNANALNLQAGGITTENGTDGIRGLVLSLKNGAATDF 286
 QY 184 NFTNTOGEI-----VN-----DFT-----L 199
 Db 287 NVTGSLGSLNKGIIIEINTVAINQGLIANAGPANAIVGTNNGAGRAAGFVSVDNKKAFTI 346
 QY 200 DAPILPKDLHPDWNYLQIKILPNDVNTAVPMPVGRVSGT-----NADD 245
 Db 347 DGQVYAKD-----MVIOSANANGQVNEFHIV-DVG-IDGTTAFKTAISTIVATONSNF 397
 QY 246 GMEFDGCGN--GQITNTDPIAQTFTTDNONPSTFNSGAMPGAN-----RYDSQLVNKH 297
 Db 398 GTTDFGNLAQVYTPDTHLNGFTGDNANNGNTAGVITFAANGTLASASADANVAVTN 457
 QY 298 IKTSFOLD-----ERINTNSNRIGRRNN-----NSKFEVIGV----- 329
 Db 458 I-TAIRASGVGVQVLSGTHTAELRGNAGSVFKLADGVINGKNQVTLVGVYLAAGIT 516
 QY 330 -DNPAHY-----IRFTDGTG-----FNF-----TN 349
 Db 517 LDGSATITGIGNGGGAALOSITTLANDATKTLTGAGANIISANGCTINFANGCTIKLT 576
 QY 350 QTOGEIVNDF-----IIDAPILPKDLHPDWNYLQIKILPNDVNTAVPMPVGR 399
 Db 577 STONNIYVDDDLAIADQGVVDASSL-----TNAQTLTISGTIGI 617
 QY 400 VSGTNADDGMEFDGCGNQTITNDPIAQTFTTDNONPSTFNSGAMPGANR--DS 452
 Db 618 IGANNTILGQFNIGS-----SKTLNGNVAINELYIANGNSVQFAHNTYLLTR 666
 QY 453 QLVNKHRIKISFQIDKEFYVPFWMTGSEENKIRLATGSLPNSERYWILDIPGTQVYLK 512
 Db 667 TTNAAGQCKIIF-----NPVYNNNTTLAAGT-----NLGSAANPLAETIN 705
 QY 513 EDSVAVFRLYNSVNSLSIFIGDSIYIFGT-----SELPJLMYSEFPRLSQTLNQV 566
 Db 706 FGSKARADYVN-----VGEGVNIYATITTTDANVGSFVNGAKNIYSGIYGGQ- 757
 QY 567 KTDIDIASSTNGTIT-----NGTT-----TADTSSGSTGAGTNTIN 605
 Db 758 QGNKFEVVALDGGTYVKFKLGNAETENGNTTIANSTQISGVYADFASADGIGIYEVN 817
 QY 606 TSGTYSNPPLNTYRSGIDSKPTSAKIDETINMADPNVIEKRIAYEIRLQIONEPITNA 665
 Db 818 TGPI--NVTLN-----KQAVPVNAKQIIVSGPGV-----VYNE--IGNA 854
 QY 666 GNF--IRNTI-----GVCGETSGSRVYLR-ASYNGDQRPYGNQFPLVY 707
 Db 855 GNVHANTDIIAFENSSLAGVLELPBGIFPNDAGNTIPLTKSYGVNTAGCFSPSVIV 914
 QY 708 FG-----YGYOQTRTG-----TFWYGYKLLNSPYDLDPVY- 742
 Db 915 SGVDSVIADQVIGDQNNIVGLGSDNGIIVNATTLVAGIGTINNNOGTYTLGSGVNT 974

QY	743	-----GETNQFRTSLT--YPMWGILTEBEGARSEFNTFYIRAGQGPSPRSIFQ	791
Db	975	PGIYYGLGELGTGAGSKFKQVYFTTIDYNNLGNITIA-----TNTIT--NDGVYITGGTAA	10235
QY	792	SGVSDNTVEYIIQSVLGFEDGIRNNLNVGVKASSPLNSRPNDCHEMTAAT-----TYL	844
Db	1026	GGIAGCTDPD--GKITLGSVGNANVRPADIFENST-----SMIYTKANNNGTYVYL	10757
QY	845	-----RSQIGLARISGLPNOQ-----PGETHQVY---SVSPGDQSS	879
Db	1076	GNAFVGNIGSDPTFVASVRFPTGSNNAGLKNIXISQVLDFTGYNLGIVNSVILGGSTYA	11355
QY	880	IK-----NIRTFPGQMLWELFETNNKSSVYTLRLAD-----SSNPASSSES	924
Db	1136	INGKIDLTNTLTIFAGTSTW-----GNNTSIEFTTLTLAGNIGHIYIACGAYQVATT	11889
QY	925	PTSLIDYNEIGVILPLLDNSFYTVAAGNVALFFSSNPGSPSYTAVN--TFPNOLSDIA	981
Db	1190	GTTTINVD-----NANAN--FSGTQ-----TYTLIGGAGFNGLGSPN	12227
QY	982	PEGGACATYISDFMGTIQPKPEYLI-QNGFTSQYARFVYTNQSF-----LNSLYDF	10311
Db	1228	FTVYSGNSRFVNY-GLIRANADQYVITRTNNAENITVTIDTISPGGAPGVGCONVTFVNA	12868
QY	1032	TPAAGATNY-----RWVDPDGLNTNOLPLKVOYLQYLDGKXYDAKLKNNN	10777
Db	1287	TNTAAYNLLAKKSADSANVGVITVDTSAITITNAQLDAVKADIAQLG-----	13355
QY	1078	LVTFSYNNFGALPSWVYFPAIGSLGILAIITLG-----LAIGI--PLPRAORLQDK	11288
Db	1336	-----NRLGALRILGTFPEWYSEBGAIPAAVGAAGDEAVDNVAVGIMAKPYTDAHQSK	13899
QY	1129	GFKTFKK-----VDTLF-----AANGSYKKIITQGTANVKKRPALGAKSKSDK	11737
Db	1390	GGLAGYAKKTYTGIVIGDFTLANNNLMIGAALG-----TYKT-DIKHODY-----KKGDK	14377

RESULT 9

OMP_B_RICCN STANDARD: PRT: 1655 AA.

AC O9KKA3; Q9KK98; Q9XC45;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Outer membrane protein B precursor (168 kDa surface-layer protein) (Surface protein antigen) (Cell surface antigen 5) (Scas) (Tromp) (Tomp B) [Contains: 120 kDa surface-exposed protein (Surface protein antigen) (120 kDa outer membrane protein ompb); 32 kDa beta peptide].

DE OMPB OR RC1085.

GN Rickettsia conorii.

OS Rickettsia conorii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsiinae; Rickettsia.

OC NCBI_TaxID=761;

OX

RN

RP SEQUENCE FROM N.A.

RC STRAIN=Malish 7;

RX MEDLINE=21442074; Pubmed=1157893;

RA Ogata H., Audic S., Remesto-Audiffren P., Fournier P.-E., Barde V., Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M., Raoult D.;

RA "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";

RT Science 293:2093-2098(2001).

RL

RN

RP SEQUENCE OF 33-1649 FROM N.A.

RC STRAIN=Indian tick cynpus, and Malish 7;

RX MEDLINE=20393643; Pubmed=10939649;

RA Roux V., Raoult D.;

RT "Phylogenetic analysis of members of the genus Rickettsia using the gene coding the outer-membrane protein romp (ompb).";

RL Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).

RN

RP SEQUENCE OF 353-1655 FROM N.A.

Query Match	3.2%	Score 204	DB 1	Length 1655
Best Local Similarity	19.5%	Pred. No. 0.0054		
Matches	272	Conservative 172	Mismatches 484	Indels 468
			Gaps	71
15 GLAVFGALGSAFEGFKOSDKNDNTQLVNQARTLDANSVRLAGLQNGSLFNTRYLRDYYD	74			
272 GRVLISKNAATDFNVTGSLGNGLKGIIEFNVAVNQOLKAMGANAIVIGT	325			
75 NFITANSTIIKLDP-----SFTKPLYGDL-----SDCGGKKVQVQIVDYDTSRNRFPQR	125			
326 NGACRAGEFVSVNDGKYATIDGQYAKADMYQISANNAVQVNFRIIVVGTGDTTAAFTTA	385			
126 QTRAVYVALV-----DEAVVHLKRI--NT--NSNRIGNRNNNSKEFVIGVDNPARYI	174			
386 ASKY--AITQNSNGTTFDGLAQAIIYPRNTMLNGNTGASNSNG-----NTAGVI	435			
175 RFTDQGEKFNFTNOTGEIVNDFILADPILPRDLHPMYNLXYIQKILIPNDVNTAVVMP	234			
436 TFDANGTLASADANVAVTNNI-----TPIEASG	465			
235 VG--RVSGTNADD-----GMPDQNGQIINTPDIATQTKTTDQNTQNSSTNSGAMPAN	285			
466 AGVQQLSTHAAELRLGAGSVFKLADSTVIN-----GKVNQTLAVGLAGLACT	514			

QY 286 NRYDSQNLVKHRIKTSFOLDERINTNSRIGN-----RNNNSK-FVIGG---V 329
 515 ITLDGSAFI-----TGDIGNAGGAALOGITLADATKITLIGGANMI 557
 QY 330 DNPAHVIFETDDGKRFNFNTOTGEIYVNF-----ILDAPILPRDLHPDMYNLXI 379
 558 GANGETINFOANGTITKITS-TONNIYVDDDLAIATDQTSVVDASLL-----603
 QY 380 QKRIIPNDVNTAVWPVWVSGTNADGMEFGCNGQITMD-PIAOTKTTEDNQNPSTF 438
 604 -----TNAOFTLIGKIGTVGANKKTLGQFNISSTKTVISDQVAINELVIG-----650
 QY 439 NSGAMGANNRR--DSQNLVKHRIKTSFOLDEKEFVPEWTEGSEENKITLATGSLPSNE 496
 651 NGAOVPAHNTYILITRTNMAOGKIIF-----NPVNNNTTLATGT-----692
 QY 497 RYWIIDIPGTPOVTLKEDSVNFSLYLINSVNSLSFIDSIYIFGT-----SELPSLWY 550
 693 -----NLGSAFNPLAEINFGSKGAAN-VDIVLVNKGVMILATNTTTDANVGSFIF 743
 QY 551 YSEPTRLSDLTALNOYKTDIDIASSTNGTTF-----NGTTF-----TA 589
 744 NAGGTNIYSGTVGGQ-QGKNFNTVALDNGTIVKFLGNATFNGNTTIAANSTLQIGANTYA 802
 QY 590 DTSSGSTAGTGTNTTTSQ-TVSNPPLNTYRSFGIDSKPTSSANKIDETNMADPNVIEARI 648
 803 DEVASADGTGIVEFVNTGPITV--TLN-----KQAPVNALKQITVSGPQNV-----847
 QY 649 YAEYRLGIONEIPITPNAGNF--IRNTIG-----GNGFTSGSRVYLAR-AS 690
 848 -----VINE--IGNNGNYHGAVTDIARENSLGAAYFLPRGIDPFNDAGRKIPITIKS 898
 QY 691 YNGDQPTGNFQPLIYVFG-----YLYGQOTRTG-----TFWGYTYKL 728
 899 TVGNKTATGDFVPSYIVLGVDSYADQVIGDDONNIVGLGSDNDIYNAATLYVAGIGCT 958
 QY 729 LNNSPDYLDSPRV-----GTETNOFRISLYTPVWAGYLETGARGSFSTPTRY 776
 959 INNOCITVTLGSGIENTPGTYVGLGTGIGAKGQVFT-----IDYNNLGN 1005
 QY 777 IRAQDPT-PESRSITQSGSDMTYEXIQSVLGFPG-----IRNNLNV-----GV--KAS 823
 1006 IITNTKINDGVYVTTGGIAG-----IGFOCKITLGVSNNGNVRVVDGILSHSTS 1056
 QY 824 FLNSRPNPNGL-----EMIAATYILRSQIGLARTSGLPNQO-PFGTTH 866
 1057 MIGTTKAN-NGVTYLGNAFVGNIGSDPTPVASVRFSGDAGLQGNITVQVDFGTYN 1115
 QY 867 QVLS--VSPGDQFSSIK-----NIRTFPGNOLWYFLFTENNKSQVYTLALADSN 916
 1116 LGISNSNVILGGTTALNKNINRTNLTFFASTGTSW-----GNNTSIEETTLTANGN- 1168
 QY 917 PDASSSFPSLIDVNEIGVILPLDLSNRYTYAAGNAL-----FSSNPGSGSYTAVN 971
 1169 -----IGNIV--ILBQAOVNATTTTITIKYQDANANANFSGTQYTLIQ 1210
 QY 972 ---TENONSLIAFEGSGAKYTSDFMCTIOFKPDEYLI-----QNGFTSOYARN-----1017
 1211 GGAENGTLGGPNEVYTGNSRFVNY-GLIRAAQODYVITRTNAEYVYNDIANSSFGA 1269
 QY 1018 -----FV--TNQSFNLVDFTPANAGTNY--RYVYDPPGNTLNONLPLKVOIOYL 1064
 1270 PGVGONVTTEVATNTAAYNNMLLAKNSANSANVGALVITDSALTINQOLDVAKIOAQ 1329
 QY 1065 DGRYIDAKIKNNNLVTFSYNNFGALPSPVWP-7A--IGSTLGIAMITIG-----LAI 1115
 1330 LG-----NRGLARLYGTPELEMAGPBEGALPAVAAGDAVDVNAV 1372
 QY 1116 GI--PLRAQRKLQDKGEKTFYFK-----VQTLT-----AAGVSYKKIITOTAN 1157
 1373 GIWAKPFYTDHAKSKKGLAGYKAKTGTGVIGLDTLADNLMIGAALG-----ITKT-D 1425

QY 1158 VKKPPALGAGKSGDK 1173
 1426 IKHODY-----KKGDK 1436
 Db
 RESULT 10
 YS89_CABEEL STANDARD: PRT: 3178 AA.
 ID YS89_CABEEL
 AC 009624; 009625; 0969D4;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein ZK945.9 in chromosome II.
 GN ZK945.9/ZK945.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peleoderinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Wilkinson-Sproat J.;
 RL Submitted (Feb-1995) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RA Durbin R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Contains 1 GPs domain.
 CC -1- SIMILARITY: Contains 1 PLAT domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z48544; CAB70192.1;
 CC EMBL: Z48582; CAB70192.1; JOINED.
 CC EMBL: Z48582; CAB70201.1; JOINED.
 CC EMBL: Z48544; CAB70201.1; JOINED.
 CC WormPep: ZK945.9; CE23697.
 CC InterPro: IPR002111; Cat_channel_TripL.
 CC InterPro: IPR005821; Ion_trans.
 CC InterPro: IPR001024; Lipoxxygenase_LH2.
 CC InterPro: IPR003915; PKD_2.
 CC InterPro: IPR00203; PKD_cys-rich.
 CC InterPro: PF01825; GPs_1.
 CC Pfam: PF00520; Ion_trans_1.
 CC Pfam: PF01477; PLAT_1.
 CC PRINTS: PR01433; POLYCYSTIN2.
 CC SMART: SM00303; GPs_1.
 CC SMART: SM00308; LH2_1.
 CC PROSITE: PS50095; PLAT_1.
 CC DR Hypothetical protein: Transmembrane.
 CC KW TRANSMEM 13 30
 FT TRANSMEM 51 73 POTENTIAL.
 FT TRANSMEM 2139 2161 POTENTIAL.
 FT TRANSMEM 2348 2367 POTENTIAL.
 FT TRANSMEM 2390 2412 POTENTIAL.
 FT TRANSMEM 2451 2468 POTENTIAL.
 FT TRANSMEM 2483 2505 POTENTIAL.
 FT TRANSMEM 2567 2589 POTENTIAL.
 FT TRANSMEM 2836 2858 POTENTIAL.
 FT TRANSMEM 2939 2961 POTENTIAL.
 FT TRANSMEM 2976 2998 POTENTIAL.
 FT TRANSMEM 3038 3060 POTENTIAL.
 FT TRANSMEM 266 1196 SER/TIR-RICH.
 FT DOMAIN 1105 1241 GLY/SER-RICH.
 FT DOMAIN 2071 1210 GPs.
 FT DOMAIN 2182 2305 PLAT.


```

Db 188 APDTASTLKVGDEGLDSTTTATPNAEP--NGAIAIAPHLRTQDGIKATSEPNWT 245
OY 213 ---NLXYORK-----ILPDVN-TAVPEPVRGVSCTNADGMFDC 250
OY 246 PESYDLISYNNMNASSTYKGAENVDAIYRYSILDSSSTAVLAELVSRITGVDLEKTYLER 305
Db 251 GNGQITNDPIAOKRTTNDON-----PSTFNGSGMGA-----NNRYDSOLANKHRJK 299
OY 306 GE-SYTFSHP--TKVNAANSNITYYDTISLANSNTGALKFSANDVSTIIIVPAQIN 361
Db 300 TSEFOLDERINTNSRIGNRNNNSKRVIGCDNPAHVRFPDDGKNEFTQTOGEIYNDF 359
OY 362 TT-----RYTESGKV-LATYGLQTIAGQVYTPSSVRVFGYDVATTTAAVGPYPKG 415
Db 360 IIDAPILPKDLHPDMNLYORKILPDVNTAVY-----PWPVGRVSGTNADGDMFDC 413
OY 416 VYLAGTVOKDT---VOYKVIIEIEND-QAVLKFYLDPTYKGEVDMRG-----460
Db 414 NGQITNDPIAOKRTTNDONPSTFNGSGMGAANKRYSOLNVKHKRISFOLD-----467
OY 461 ----TDTGTIELLT-----SPTTYKVGTI-----YD--YVNSKITAPFTIDPTKNM 504
Db 468 -----EKF-VYEPWTSSEENKI-----TRLATGSLPSNERWILDPGTPQ 508
OY 505 VFKESEONEGSKRYVIAOMSGDETCKGICYKIATQVWTKLGTNMGWF-----556
Db 509 VTKEDSVNFSRLYNSVLSIFGDSIYITGTSLSPLWYYSFPTSLDLTALNOYKT 568
OY 557 -DYDDDAQI-----KFNKNGP-----WPAGVONLRNMTPTAVET 592
Db 569 DDIKASSDNDGT-----TNG--TTTADTSSGSGTAGTGNTN---TSQTSNPLN 616
OY 593 TYIKESSKYGDVIEYDIDGKQIVNSVDTPKSAIGEYNDVRRPASIIVADGVY 652
OY 617 TYRFGIDSKPTSAKIDETNMADPNVIEARIVAEVYLGIONELPITNAGNFRNTIGV 676
Db 653 FKKEKSDSAKTTGGVAVGTTVK-----YIEKAGSVN-----V 687
OY 677 GFTSGSRVY-LRASYNDDORPTGNFOPFLYVGYLQOORTGCTFWGYTKLNNSPD 735
Db 688 NFWDINGKVIKAPVDEDAKGVNYDTDL-----DOKLASTFEGKEKIVPADDP 740
OY 736 VLDSPVGTETNOFRRTSLITVPMGGLTEGASFSNTPYIRAGDTPSRISIFOSGS 795
Db 741 V--GKVG-----KGNLIEVG--NMT--AKGIDPTGKIEAGVKN 774
OY 796 DNTYEYI--QSVLGFDEIRNLL-----NVGKASFSNMRPN-----831
Db 775 EVIYVRAVGSVVVYKDEGVNIDKPEIDVSDAPVDDATYTTDKKNEIITKDSRYV 834
OY 832 -----PNG--LEKIAATYLRISOIG--LARTSGLPNOQ-----PFGTT--HOVI 869
Db 835 LVPKSTDEENGKVIETITVIVYOKVANNMPEIPNVETDRPKVPYFPTEPDEPID 894
OY 870 SVSPGDDFSSIKIRITFPGNOLMYLFTNENKSSVYLRKLADSSNPASSFSPTSLI 929
Db 895 PTTROTIN-GEVPIPIPV-PG-----YT-----PYDPDNTPLAPI 927
OY 930 DVNEIGVILPLDINSFTVNAAGVAFSSNPG--SPGSYTAV--NTENONISDLA- 981
Db 928 DNDPG-----KGVYPTPENPGVDPIPVYPAKVTNHNHVEGPNPIAP 972
OY 982 -----FEGSAGKYSDFWGTIOKPEEYILONGCFISOVARNFVNO--SFLNS 1027
Db 973 QEGGTRKPKNSIGYEFTGKTVTEDEGNTTHLYKTPPEVKN--TYVNVYTTDEGVKE 1029
OY 1028 LVDETPAN-AGTNYRVVDPDGNLTNOMLPLKV-----QIOYLDGKYDAKLNNLY 1079
Db 1030 PYVDPTSPGTPYD-----TTDKKPTITFKGEYELVAVDGTENGKVESETV 1080
OY 1080 TFSYNNFGALPSSVVPALGSLGLAIIILGLAIGIPLAAOK-----LQDGKFTPF 1134
Db 1081 TYVYRKVEYTPAKKVVVNHVDEE-----GNVAPQDEGTRKPKNSIGYEFTG 1126

```

```

OY 1135 KKVDTLTAAGSVYKKIITQITANVKKRPAALGAGSGDKKPAAPAKP-SAPKAS 1193
Db 1127 KVTYDEGNTTHYKK-----TPAKKVYTNHVEDEGN-----PIAQEGGTTPKRO 1172
OY 1194 SPA-----KPGKSGAPTRKPAKPAKPPAP 1222
Db 1173 ISGEYVRFVVEDEGNTTHYKLSKNKPTTPKTPAKPOAGKVASGAOLP 1224

RESULT 12
P3P_LALCIC STANDARD: PRT; 1902 AA.
ID P3P_LALCIC
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE P11I-type proteinase precursor (EC 3.4.21.96) (lactocoeplin) (Cell wall-associated serine proteinase).
DE PRP.
GN Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OS Plasmid.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_Taxid=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC SRRAIN-SK11;
RX MEDLINE=89340435; PubMed=2760036;
RA Vos F., Simons G., Slezan R.J., de Vos W.M.;
RT "Primary structure and organization of the gene for a procaryotic,
  cell envelope-located serine proteinase."
  J. Biol. Chem. 264:13579-13585(1989).
RL [1]
CC -1- FUNCTION: PROTEINASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
  GROWTH OF THE BACTERIA ON MILK.
CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
  specificity, although some subsite preference have been noted,
  e.g. large hydrophobic residues in the p1 and p4 positions, and
  pro in the p2 position. Best known for its action on caseins,
  although it has been shown to hydrolyze hemoglobin and oxidized
  insulin B-chain.
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
  an amide bond (potential).
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib.ch).
CC -----
CC EMBL: J04962; AA003533.1; ALT_SEQ.
CC HSSP: P00782; 2SBR.
CC MEROPS: S08.019; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR006192; LPXTG.
DR InterPro: IPR003137; PA.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR Pfam: PF02225; PA.1.
DR Pfam: PF00082; Peptidase_S8.1.
DR PRINTS: PR00723; SUBTILISIN.
DR TIGRfam: TIGR01167; LPXTG_anchor.1.
DR PROSITE: PS50847; GRAM_POS_ANCHORING.1.
DR PROSITE: PS00137; SUBTILASE_ASP.1.
DR PROSITE: PS00136; SUBTILASE_HIS.1.
DR PROSITE: PS00136; SUBTILASE_SER.1.
KW Hydrolyase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW Signal; Plasmid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT P11I-TYPE PROTEINASE.

```


FT PROPER 1871 1902 REMOVED BY SORTASE (POTENTIAL).
 FT ACT SITE 217 217 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 1867 1871 LPXG SORTING SIGNAL (POTENTIAL).
 FT MOD.RES 1870 1870 AMIDE-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 1902 AA: 200550 MW: 87CEBAA9345F9D3 CRC64;

Query Match 3.1%; Score 199; DB 1; Length 1902;
 Best Local Similarity 18.8%; Pred. No. 0.012;
 Matches 315; Conservative 162; Mismatches 529; Indels 672; Gaps 74;

QY 20 GALGSAFEGKQSDKSDNTQVNOARTLDANSVRLAGLQNGSEFTVLRVDNFIITA 79
 DB 385 GTSAGTEGVNKKDYGLQDNEMVSGPST-----SRGATVASAENT---DVITQAVTI 434
 QY 80 ANGIIKIDSTFKPLVGLDLS-----DDCGYKVKQVSVYTTSRN----- 120
 DB 435 TGTGTGLQGPETIQSSHDFGTSFDQKFTYIVKDAAGNLKGLADYTADAKGKIAIVKR 494
 QY 121 ---RFDQROTAVYA---LVNDEANVHLKRIINTSNRIGNNNNSKFEVIGVDNPAHV 173
 DB 495 GEFSFDKQKVAQAAGALIT-----VNTD-----CTATPMIS 528
 QY 174 IFTDDGTGFNTNOTGEBIVDEIILADILPKDLHP-DWYNLYIQRLILPNDVNTAVVP 232
 DB 529 IALTTFPFEGLSVYV-GQKLVDMV-----TAPDDSLGKVTILAMLPQKKT- 575
 QY 233 WVGVRSGTNADDGMDGNGOITNTDPIAOTK-----TTTONONSTPNSG- 279
 DB 576 -----EDKMSD-----FTSYGPVSNLSFKPDITAPGGINWSTONNGYTMMSGT 619
 QY 280 -----AMPANN---RYDSQNLVHKRIKTSFOLDRIKTSNRIGNRNN 320
 DB 620 SMASPIIAGSQAALLKQALNNKNPNFYAYKKOL-KOTALDPLKTYEMNT-AQPIINDIV 676
 QY 321 NSEKV-----IGVD-----NPAHVI-----RFTDGTGFNTNOT 351
 DB 677 NNVTISPRQAGLVADVKAIDAILEKPNSTVAENGYPAVELKDEFTDKTEFLKPTNRT 736
 QY 352 QGEIV-----NDFILDAPI----- 365
 DB 737 THELTYQMSDNTDTNAVYTSATDPNSGVILDKKIDGAIKAGSNTVPAGKTAQIEFTIS 796
 QY 366 LPKDLHP-----DWYNLYIQRLI-----LPNDVNTAVV 393
 DB 797 LKPSFDQGFVEGFLNFKGSDSRLLPYMGFFGDNNDKIKYDLSLNGITVSPAGNFQIV 856
 QY 394 PMPVGVSGTNADMGHFCGNGOITNTD---PIAOTKTTTND-----QNPSTFNSG 441
 DB 857 PLKKNKNTCTOYGGAVTDADGKIKTVDDOALIAFSSSDKNALYNDISMKYLLRNISNVQVD 916
 QY 442 AMPGANRRYDSQNLVHKRIKTSFOL-DEKFFVY---PEWNGS---EENKNITRLATGSL- 492
 DB 917 ILDGQGNKAVTTLSSTSNRRKTYNAHSOQYIYNAAMGQYTDQDGNKLTADDSYTY 976
 QY 493 -----PSNERWIIDI-----PGTQVTLKEDSVVVFSLYNSVNSLSFIDSIY 538
 DB 977 RISGVPEGGDKQVFEVPPFKDSKAPTVAHVHVALSATEKNGKQIVL----- 1022
 QY 539 IGTSELPLMWYSFPTRLSDLTALNQVTD-----DIEASSTNGCTTNG----- 584
 DB 1023 ---TABAKD-----DLSGIDATKSVTEINEVNLDTFTDAGTADGYKIKETPLIS 1071
 QY 585 -----TTTT- 588
 DB 1072 DEQAQALGNGNSAELYLTDNMSNATDODASVQKPGSTFEDLIVNGGIGIDKISSTTGV 1131
 QY 589 -ADTSSGSTGAGTG-----NTT-----NTSOTVSNPTL- 616
 DB 1132 EANTOGGGTYTTSQTPPAVVDGTYTDAOGKKHDLNNTYDAATNFSFASHPVNTADYAAQV 1191
 QY 617 -----TYRSF-----GIDSKPTSAKIDETNMADPNVIE--- 645

DB 1192 DLYADKAHTQLLKHFDTKVRLMAPFTTDLKFNNGSQDSTSEATIKVGTVASDPTKRVNCH 1251
 QY 646 --ARIYAEIRLQNIPIPTMAGNFIKNTIGVGFSTGSRVLRASVNGDQRPIGNFOP 703
 DB 1252 TVAAALDQHHFVS--DVPVNYGDNTIKVATDKDGNTEOKRITSSYPDM----- 1301
 QY 704 FLYVFCGLYQOQRTGTFFWYGYKYLNNSP--YDVLDSPRVG--FTENOFRRTSLTPYM 759
 DB 1302 -----LKSXTFDDGVKFGNKKFNATSAKFTD-----FKTGATITGKVKHPTTLQVD 1350
 QY 760 GGYLIEGARFSNT-----PYRAOGDTPESRSIFQSGYSDNTYEYIQSVLGFDDG 810
 DB 1351 GKQIPKIDDLTFSEFTLDLGLGOKPPGVVGGDTQKTK-FQ----- 1390
 QY 811 IRNNLVGVKASSFLNSNPNRNGLEMIATYILRSQIGLARISGLPNOQPGT---THQ 867
 DB 1391 -----EALSF-----LDVAVPTLSLSDSTDPAPYTTNDPNQITGTATDANMO 1432
 QY 868 VISV-----SPGDQESSIKNIRTIFFGNOLWFLFENNNK-----SSVYTLRLADSSND 918
 DB 1433 YLSLSINGSSVASQYEDI-NINSGKRGH-----MAIDQVYKLLGKNAVLTVAVTSEDMT 1486
 QY 919 ASSSF-----SPTSLIDVNEIGVILPILDNSFTVNA--GNVALSSNPGSFG 965
 DB 1487 TTKNITVYVEPKKTLAAPTVPSTETPEAQTVTL-----TANNAATGETVOYSADG- 1537
 QY 966 SYTAVNTFNONLSIDIAFEGSGAKYTSDFMGTTIOFKPD-----EYLIONGFTISOYA 1015
 DB 1538 -----KTYQDV--PAAGYITAN--GTRFKSTDLXGNEPAVDVYVNTIKADDPRA 1584
 QY 1016 RNEVTNOSFLNSLVDTTPANAGTNY-----RVVYDPDG---NLTNQNLPT 1056
 DB 1585 QLOAAQELNLNLASAKTSLTASGKYDATTTLAATAQKQATRLATDNTASVSLTGAND 1644
 QY 1057 LKVOIYLDGKY-YDAKLKNNLV-----TFSTNNFGALPSWV 1094
 DB 1645 LQTAIQLQALPADKETSILNLOSVKDALGTDLGNQDTPSTGKFTT---AALDDLVA 1700
 QY 1095 PTAIG-----STGIILAIMIILGLAIGIPLARAQRIKDDGKFTTPK---VDTLT 1141
 DB 1701 QMOAGTQTDQLOATLAKLIDLEVLAKEIKATPAEYGNAKDAATGKTWADLADTL 1760
 QY 1142 AAVGSYVKKIITOTAVNKKRPPALGAGSGDKGAPAAAPAPAPKPS----- 1188
 DB 1761 SGQAS-----AASDKLAHLQALQSLKRTVAVAEAKVYVGGDGTGSDKGGQ 1811
 QY 1189 ---APKASSPAKPTG-----PKSGA--PKRP-----TAPKPAAPK 1218
 DB 1812 GTPAPAPGDTGKDKGDEGSGQSSGGINIPTRPATTTSTTDDTTDRNGOLTSKGLAPK 1869

RESULT 13

OMPA_RICRI STANDARD: PRT; 2249 AA.
 ID OMPA_RICRI
 AC P15921;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Outer membrane protein A precursor (190 kDa antigen) (cell surface antigen) (iOmpA) (iOmp A).
 GN OMPA.
 OS Rickettsia rickettsii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales; Rickettsiaceae; Rickettsieae; Rickettsia.
 OX NCBI_TaxId=783;
 OX RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-R:
 RX MEDLINE=90354033; PubMed=2117568;
 RA Anderson B.E., McDonald G.A., Jones D.C., Regnery R.L.;
 RT "A protective protein antigen of Rickettsia rickettsii has tandemly repeated, near-identical sequences.";

OY	357	NOFILDAPILEKDLHPDWYNLQIKILLPNDVNTAVPMPFVGVSCT--NADDGMD-CGN	414
Dd	1136	-----PVV-----	VTGADINTGANNNGIYTFNGN 1159
OY	415	----GOITNDPIACTKTITTDNONSPFENSGAMGANNRYDSOLJNKHRIKTSFOLDEKE	470
Dd	1160	SVTGTGIGIGTNNMLA-----TVNAGACITIQAGGSLAANN-----IDFGARSTLEFNGPLD---	1209
OY	471	VYEWEMTGESEBKNTITLATGSLPSNERWYLLDPGTFOYVLKEDSVWFERLYLANSVNL	530
Dd	1210	-----GGKAPIYEFKGI-ANGNALLINV--NTKLLTASHLITITVAEINIGANGJF	1259
OY	531	SE---IGDSIYIFGTSELSPLMWYSPPT-----LSDLTAL--NOYKTDIDFASSTONG	579
Dd	1260	TIDASVGD-VTILMAONI-----NFRADSVLVLSNLTVGVANNILLADIVAPGADEG	1312
OY	580	TTT-NGTTTADTSSG-----STGAGTGNNT-----SOTSPRLNTY	618
Dd	1313	TVVNGGVNGLNNGSNVAGTARNIGGGGNKNTLLIYNAVTTDDVNLBEIQVNLKN	1372
OY	619	RSFGIDSKPISAN-----KIDETMADPNVIERIYAERYLQIONEIPITNAGNFRNTIG	674
Dd	1373	ADF-----FSTTFNNGAIQINDATY-----TIDA-----NNGN-LNIPAG	1407
OY	675	GVGFTSGTSGVILRASVNGDQRT--GNFOPLVYGYL-----GYOQRTG--TFWY	723
Dd	1408	NIOFMAHQALVILQNSSGNDRTITLGNIDIPDNDGVIYILNSVYAGKKLTLAGKPTG-	1466
OY	724	GYVLLANSPIVDVDSPRVGTETNOQFRTSLTFYPMVGYLTEGARS-----FSMTPIR	778
Dd	1467	GAHKIOTILFKGAGDCSTAGT--TFMTNLVIDITG--OLEIGATTAVVLENOAVOLT	1521
OY	779	AGGDPEESKSLFOSGYSDNTYERYIQSVLFGDGIKNNLNVKFKSSPLANSRPNPGLXMI	838
Dd	1522	QCGNI-----GGFLD-----FNAGKAVTLLNNNVN--AGAVOINQGTN-NGTLIV	1564
OY	839	AATTYLRSGIIGLARTSGL-PNOQFETTHQVIVSPPGQFSSIKNIRTIPPGQLMW---YF	895
Dd	1565	LGASNLNRXVNGIAMLK-----VQAGVVTIAKGGKVKIGIEIOGTG---NTLLPAPH	1613
OY	896	LFTNENKRSVYTLRLADSSNPDASSFSPSLIDVNEIGVILPLDNSF-VTVNAAGV	954
Dd	1614	NLTGSIKKTGGQALKI-NFMNGSGVSGVYGA--ANSVGDITTAGATSPASSVNAKGTG	1669
OY	955	AL-----FSSNPGSPGSYT--AVNTFONLSDIAFEGSAAKTSDEMGTOIEKPEVY	1005
Dd	1670	TLGTTTFATFTINTGAVTLAKGSLTSFAKNTATSFVANS--TINSLSLAFNSN--I	1725
OY	1006	IONGFTSOVARNFYT--NOSFLNSLY-----	1029
Dd	1726	TGGGTTLTILGANOVTYTGSEFDTLTLNTTDDGAAKSGNLLIKGSTLDSGVSTIAL	1785
OY	1030	-----DFEPANAGTVRYV-VDPDGNL--TNONLPLKVOYQVLDKYYDADAKLKNNL	1078
Dd	1786	VVTATNFMNMNISPTKTYVJISAETAGGLKPMRSKENVKTTIN-----NDNR	1832
OY	1079	VTFSYNNFGALPSWVYPAIGSTGI-----LAIMILG--LAIGIPLR-----	1120
Dd	1833	VDFETD-----ASTLTLEADLIADVLDGOFAPGAPLANIPNANIKSLE	1878
OY	1121	-----AQRKLDKGFTTERKVDVTLTAAYGSYKKIITQF-----NVKKPK	1162
Dd	1879	LMEDAPNGSDAQAQANNFGMLPTLOEADATHTLIQDVYKPSDTIAAVNNQVYASNISSI	1938
OY	1163	AALGAG---KSGDKKPAAPAKPAAPAK	1186
Dd	1939	TALNARMKDVOSGNKGVPSSGDEDMAK	1966

01-OCT-1993 (Rel. 27, Created)
01-OCT-1993 (Rel. 27, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
MSB2 protein (Multicopy suppression of a budding defect 2).
MSB2 OR YGR014W.
Saccharomyces cerevisiae (Baker's yeast).
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932.
[1]
SEQUENCE FROM N.A.
MEDLINE=92383951; Pubmed=1514328;
Bender A., Pringle J.R.;
"A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence
defect".
RL Yeast 8:315-323(1992).
[2]
SEQUENCE FROM N.A.
STRAIN=5288c;
MEDLINE=97435481; Pubmed=9290212;
Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
"Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
chromosome VII".
Yeast 13:1077-1090(1997).
-1- PFM: O-GLYCOSYLATED IN THE SER/THR-RICH REGIONS (PROBABLE).
CC -1- SIMILARITY: SOME, TO YEAST HKR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: M77354; AAA34798.1; -;
DR EMBL: 272799; CA96997.1; -;
DR PIR: S25370; S25370.
DR SGD: S0003246; MSB2.
DR GO: GO:0005887; C: integral to plasma membrane; IMP.
DR GO: GO:0005034; F: osmosensor activity; IMP.
DR GO: GO:0000283; P: establishment of cell polarity (sensu Sacch. .; IGI.
DR GO: GO:0006970; P: response to osmotic stress; IMP.
KM Transmembrane; Glycoprotein; Repeat.
FT DOMAIN 698 714 7 x 17 AA TANDEM REPEATS.
FT REPEAT 698 714 1.
FT REPEAT 715 731 2.
FT REPEAT 732 748 3.
FT REPEAT 749 765 4.
FT REPEAT 766 782 5.
FT REPEAT 783 799 6.
FT REPEAT 800 816 7.
SQ SEQUENCE 1306 AA; 133114 MW; 67D5D984D5CA4A6D CRC64;
Query Match 3.1%; Score 197; DB 1; Length 1306;
Best Local Similarity 18.5%; Pred. No. 0.0091;
Matches 234; Conservative 176; Mismatches 456; Indels 396; Gaps 56;
QY 12 LIGLAVFGLGSNS-FGF-----KSDKSNNDTQLVNARTILDANSVRLA----- 56
DB 8 LLSLTVLSGLSARASPDFFFGNGTQDOAQOSOGVSFTNEA-SQDSSTTSLVYATSQ 66
QY 57 GL--GONGSLFNTVLRVNDNFITAANGTIIKIDSFTKPLXGLDLSDDGGRVYKQVSD 114
DB 67 GVSHSGATIVSATISSLPSTWYDASS-----TSQTSVYASQESD---YAVNQ--NS 114
QY 115 YTSRNRFDOQRTRAYVALLVNDEANYHLKRTNSNRIGNRNNSKFVIGVDNPAHVI 174
DB 115 WSASTNQLPSTSTSYAPFSTADPAASVNAAS-----DVSTASVPIDTSANSI 166
QY 175 RPTDDGKRFNTQDGEIYNDPILAPILPKDLHPWYMLYIQRKILPMDVNTAVVPM 234
DB 167 PFT--TTSNIETTSAPLSD-----TP 187

235 VGRVSGTNADDEGFCDCNGOITNTDPIAOKTTTDNON--PSTFNSGAMPGANRVDSQL 292
DB 188 LISTMSADANVFSSAN-----PISASLTITDSSESDOTSTGALPVGSS----- 234
QY 293 NVKHKIKTSFOLDERINTNSNRIGNRNNSKFVIGVDNPAHVIREFTDGKFNFTNQ 352
DB 235 -----ADRSSSEILVQSSA-----DFSSPS 256
QY 353 GEIYNDPILDAPILPDLHPDWYMLYIQRKILPMDVNTAVVPMVGR--YSGTNADGAF 410
DB 257 PTTTDSLAAPIQTS-----SSFTTSALPLVSTDDGSSAPVPS 301
QY 411 DCGNGOITNDPIAOKTTTDNONPS-TFN--SGAMPGAN-NRVDSQLNVKHKIKTSFQL 466
DB 302 MSAAGQIAS-----STDNTMSETSLTETVDGSSVSTVALSAPLQITSTN 354
QY 467 DEKFYVP--EMTGSSEKNITRLATGSLPSNERVWILDIPGTPVTLKEDSVNVFSRLYL 524
DB 355 SFSIVSPSVSFVPSQSSSDVASSSTANVSSS--FSDIP--PQSTSGSVSV-----A 404
QY 525 NSVNSLSFICDSTIYICGTSRLPLWYRFPFTRLSDLTALNQVKTDDIEASTONG--TJT 582
DB 405 QASALAF-QSSTREYVGA-----ASTMSLSLSTT 434
QY 583 NGTTTADTSS-GSTAGTGNTN--TSQTVSNPTL-----NTYRSFGIDSKPT-----S 629
DB 435 SLOSTLIDSSSLASSSSSDLDYGVSSASTIPLISASQASTSSFSVPSVSVFVS 494
QY 630 ANKIDETNMADPNVIEARIYAEYRLGIQNEIPTNAGNFRNTIGGVFTSGSRVLA 669
DB 495 QSSSDVASTAPSV-----VSSFSYSLQAGSSSMNPSSSTIYVS 536
QY 690 SYNGDQRPICNFPFLVFGYLGQTRTGTFWYGYTKLLNPNPYDVLDSPRGTETNOF 749
DB 537 SSTGSSEESAATAATLSG-----SSSTYVAGN--LQSQP-----PSTSSLSES 580
QY 750 RRTSLIYVVMGGLTEBGRASFGNT--PYIAQDTPESRIIPSGYSDNT--YETIOSVLG 807
DB 581 QATS-TSAVL-----ASSSVSTTSYTTAGASSTBASLISTSAETSOVSYSOS--- 629
QY 808 FDGIRNNLVNKAASFLNSRPNPGLMEIAATYLRSGIQLARTSGLPNOQFCTTHO 867
DB 630 -----TVALQTSFPASS--TTEGSE-----TSSQGFSTSV 659
QY 868 VISV--SPGDPSISIKNIRTFPGNQMLYTLFTNENKSSVTLR-----LADSSNPASS 921
DB 660 LVQMPSSISSEFSPSQTTQM-----NSASSSSQYRTISSTGILSQVSDTSVSX 707
QY 922 SFSPTSLIDVNEIGVILPLLDNSFYVNAAGNALFSPNGSPGSYTAVNTFNQNLSDIA 981
DB 708 TTSSSSVQVSDTPV-----SYTSSSS--VSQVSDTPVSTTSSSVQVSD-- 753
QY 982 FECSGAKYTSDFWGTIOFK--PDEYLQNGFTSQVARNFYTNOSFLNSL--VDFTPANAG 1037
DB 754 ---TPVSYTSSSSSVQVSDTPVSYTSSSSSVQVSDTPVSPSRSSSVQVSDTPVPST 810
QY 1038 TNRRVYVDPDGNLTNQMLPLKVOIQYLDGKYIYAKLKNLNLVFPYVYNFALSWVPPTA 1097
DB 811 SSRSSVQSTSSSQ-----PTTSSQ-----RFTJSTHGALS----- 842
QY 1098 IGSTLGLIAMIILGLAIGLPLAORKLQDKGKFTPEKVDTLAAAGSVYKKILIJOTAN 1157
DB 843 -----ESSVSQASSETTSSINATASETHSIQTAA 874
QY 1158 VKKKPALGAGKSGDKRP--AAAKAPAAPK-----PSAPKASSPAKPTGKSGAPRK 1208
DB 875 QSTTLSTFDANSSASAPLEVATSTPSPSKASSLLTLPSTSSLSQVATWTNVTSLTTE 934
QY 1209 PT 1210
DB 935 ST 936


```
QY 851 ARTSGLPNOQPEGTTHQVSVSPGQFSSIKNIRITIFPGNOLWYFLFTNENKSSVYTLR 910
Db 784 -----VTNLTLEFASGSTM-----GNNTSIETTLR 808
QY 911 LADSSNPDASSSEPTSLIDVNEIGVILFLDNSFTYVNAAGNAL-----FSSNPGSPG 965
Db 809 LANGN-----IGHIV-ILEGAQVNTTTTGTITIKVODNANANFSGTQ 849
QY 966 SYTAVN---TFNONLSDIAFEGSGAKYTSDFWGTIOFKPDEYLI-----ONGFTSQVAR 1016
Db 850 TITLIGGARFNGTLCSPNFAYTGSNREFVNY-SLIRANODIVITRTNNANENVYINDIAN 908
QY 1017 N-----FV--TNOSFLNSLYDFTPANAGTNY--RVVYDPDGNLTNQNPLK 1058
Db 909 SPFGAPGVQDVNTTFVNATNTTAAYNNLLAKANSANSANFVGAIYVDTSAITNVQDLA 968
QY 1059 VQIQYLDGKYIDAKLKNNNLVTFSYNNFGALPSWVP-TA--IGSTLGIATMIIIG--- 1112
Db 969 KDIQAQLG-----NRLGALRYLGTPTAEMAGPEAGAISSAAVAAGDEA 1011
QY 1113 ---LAIGI---PLRAQRKLODKGFKTTFK-----VDTLTP-----AAGSVYKKI 1151
Db 1012 IDNVAIGIWAAPPYTDAHQSKRGGLAGYRAKTTGVVIGLDTLANDNLMIGAATG----- 1065
QY 1152 ITQTANVKKKPPALGAGKSGDK 1173
Db 1066 ITFT-DIKHODY-----KKGDK 1081
```

Search completed: October 10, 2003, 16:33:55
Job time : 34 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 16:18:46 ; Search time 34 Seconds
(without alignments)
3462.073 Million cell updates/sec

Title: US-09-901-572A-4
Perfect score: 6413

Sequence: 1 MWISKKLSYFLIGLAVFG.....APTKEPTAPKPAKPTAPKE 1224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_76:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598.5	9.3	1052	2 C64221	hypothetical 114k
2	556.5	8.7	1218	2 J30069	hypothetical PI OP
3	501.5	7.8	1122	2 T18346	MGCI protein precu
4	337	5.3	1144	2 A36968	PI-like adhesin pr
5	328	5.1	1627	1 IJYMAP	adhesin PI, group
6	310.5	4.8	1635	2 A41480	140k adhesin precu
7	285	4.4	1444	1 A30588	outer membrane pro
8	247.5	3.9	1029	2 T30852	hypothetical prote
9	245.5	3.8	1487	2 AG2560	hypothetical prote
10	242	3.8	3194	2 D71917	toxln-like outer m
11	233.5	3.6	4199	2 S76412	hypothetical prote
12	232.5	3.6	2021	2 A97859	190-kDa cell surfa
13	225.5	3.5	4688	2 F82885	hypothetical prote
14	223.5	3.5	1999	2 AB2018	hypothetical prote
15	223	3.5	2893	2 A64556	toxln-like outer m
16	219.5	3.4	2340	2 B71704	cell surface anti
17	218.5	3.4	1428	2 AC2224	hypothetical prote
18	218	3.4	1107	2 AC0976	probable autotrans
19	216.5	3.4	1643	2 D71630	outer membrane pro
20	215	3.4	1238	2 A64596	hypothetical prote
21	214.5	3.3	1651	2 UC1340	outer membrane pro
22	214	3.3	2902	2 C71953	toxln-like outer m
23	212	3.3	1910	2 AF0394	probable adhesin h
24	211.5	3.3	3972	2 S75251	hypothetical prote
25	210	3.3	1335	2 T17508	glycoprotein Vp260
26	209.5	3.3	1939	2 D97316	probable S-layer p
27	207.5	3.2	1180	2 AD0123	probable autotrans
28	207.5	3.2	1780	2 E86719	hypothetical prote
29	207.5	3.2	1588	2 A86036	probable adhesin 2

30	207.5	3.2	1588	2 H91188	probable adhesin E
31	206	3.2	5291	2 F90696	hypothetical prote
32	205	3.2	2817	2 B97033	uncharacterized pr
33	204	3.2	1655	2 E97835	hypothetical prote
34	203.5	3.2	1377	2 D90538	hypothetical prote
35	202	3.1	5188	2 B85547	probable RTX fam1
36	201.5	3.1	3890	2 C89921	hypothetical prote
37	201	3.1	2020	2 C48399	ABC-type transport
38	201	3.1	4936	2 AH2515	hypothetical prote
39	200	3.1	1369	2 T17504	hypothetical prote
40	199.5	3.1	1844	2 D71612	hypothetical prote
41	199	3.1	1256	1 A43829	muramidase-release
42	199	3.1	1777	2 T34369	hypothetical prote
43	198.5	3.1	2249	2 A41477	190k surface anti
44	197	3.1	1306	2 S25370	MSB2 protein - yea
45	197	3.1	3624	2 AD0835	large repetitive p

ALIGNMENTS

RESULT 1

C64221 hypothetical 114k protein (Mgpa 3' region) - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 17-Nov-1995 #sequence.revision 17-Nov-1995 #text_change 07-Dec-1999

C:Accession: C64221; J00092; S18702; S18703

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.

M.; Fuhmann, J.; Nguyen, D.; Uitterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick,

C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MWID:96026346; PMID:756993

A:Accession: C64221

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1052 <TIGR>

A:Cross-references: GB:U39696; GB:U43967; NID:g1045869; PID:g1045877; TIGR:MG192

A:Experimental source: strain G-37

R:Inamine, J.M.; Loechel, S.; Collier, A.M.; Barile, M.F.; Hu, P.C.

Gene 62, 259-267, 1989

A:Title: Nucleotide sequence of the Mgpa (mgp) operon of Mycoplasma genitalium and co

A:Reference number: J00090; MWID:90060815; PMID:2583522

A:Accession: J00092

A:Molecule type: DNA

A:Residues: 1-1052 <INA>

A:Cross-references: GB:W31431; NID:g150157; PIDN:AA25421.1; PID:g150160

A:Experimental source: strain G-37 (ATCC 33530)

R:Peterson, S.N.; Schramm, N.; Hu, P.; Bott, K.F.; Hutchison III, C.A.

Nucleic Acids Res. 19, 6027-6031, 1991

A:Title: A random sequencing approach for placing markers on the physical map of Myco

A:Reference number: S18693; MWID:92051396; PMID:145886

A:Accession: S18702

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 902-964 <PET>

A:Cross-references: EMBL:X61525

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 199

C:Genetics:

A:Genetic code: SGC3

Query Match 9.3%; Score 598.5; DB 2; Length 1052;

Best Local Similarity 23.7%; Pred. No. 7,1e-24;

Matches 270; Conservative 166; Mismatches 409; Indels 295; Gaps 49;

QY 260 PIAQRT---TTDNQPSFNSGAMPQANNRDSQANVKHRITKSQQLDRITNSN--- 313

Db 21 PALANTFLVKEBKSNTAVTTPFATPTTSK--SDVLSLAQLDSSYQIADQTHNTLEV 78

QY 314 RIGNRNNNSKFGVIGVDN-----PAHYIRFTDQGTNRNFN-QTQGGIVDELI 361

Db 79 LFKSRDVKVYESSGSNNISFSDTSQGEKPSVVEFT-----NSTNIGIKTWVKKYOL 132

QY	521	RLXVNSVNSLSFIQDSIIYIFCTSELSPLMY-	-----	SEPRRLSDLT	561			
Db	471	QLYHPTSTQAQLAYLVNGQIYVWGSDBRPSLWYVGEDESGKATWAKTELWGDOKQOF	530					
QY	562	ALNOVK-+DDLEASTDNGTNTNGTNTTADTSSG-	-----	STGA-GRGNTNT	606			
Db	531	VENOLGPRDSDNSDSKSNLNKAGOLTOPAYLIAGLDVYADHLVFAAFAGAVGDMTDS	590					
QY	607	SOIVSNPLNTYRSFGIDSKPTSAKIDETWADPNVIEARIYAEVRLGIONEIPIITNAG	666					
Db	591	SASTYNOALMASTTAGLSDS-	-----	-GGYKALVENTAGLNGPI	627			
QY	667	N-----	FINTJIGVGFTSGSRVVLARASVNGORPTGNQPLVYFGYIGYQOT	716				
Db	628	NGLTLLDTFAYIYVPSVSMKSGKSONNEEOTTYPYKSDQKATAKIA	-----		673			
QY	717	RTGTFWYGYTKLNLNSP-	-----	YDY-	-----	LDSPRYGETNOFRRTSLYP	757	
Db	674	-----	SLINASPNSYGDGOVYFEDALGLNFNKLMBERLPSRTDL	-----	LYY	718		
QY	758	VMGCLTE-	-----	EGARFSN-	-----	TPYIRAGDLPES-	RS	788
Db	719	-----	GIVNSELSKARENAOSTSDSNSNTKVKWTNASHYLPVYYYS	-----	ANPEEAGNRRA	774		
QY	789	IFOSGYSDNTE-	-----	YIOSVLGFDDIRNNLVGVASSFSLNSRPNGLIEIATYT	842			
Db	775	EQRNGVKRSTLESQATOFANSNLNF-	-----	TGLKAGVDPAPVAGHKRYNSAVILLVR-	829			
QY	843	YLRSGIOLAPRTSGLPNO-	-----	PGTTHOYIVSPPGQFSSIKNR	884			
Db	830	-----	GVRRNFNDTDLKLDSTDKNSEPIFSYTPPGSAESAVDL	-----	TLTKOV-	876		
QY	885	TIFPGNOLMYELFN-	-----	ENKKS-SVYTL-FLADSSNPDASS-	SFSP	926		
Db	877	TYNESGLMETTFDNGEKPTYDGOQOQKNNKGAVITVSTRTGIEFNEDATYTTLSQAPA	936					
QY	927	SLIDVNEIGV-	-----	ILPLDLSFTYVNA-	-----	AGNVALESSNPGSPGSYTAVTENTQ	975	
Db	937	ALAVOONGIASSODDLTGILPLSDESAIVITKDOTWTGKVDIYKN-	-----	TNGLEFEKDDQJSE	993			
QY	976	N-----	LSDIAFEGSGAKYTSPPMGWIOKRPDEYLIQNOFTSOVANAFTYNOFSLNSL	1028				
Db	994	NKRRDNCGLVITYEG-	-----	YDVGWVADPANSVLOARMLTDTYDEVINNPDIJOSF	1048			
QY	1029	VDFTPANAGTYRVAVVPDGNLJTNQNLPLKAOIQLDKEY-	-----	DAKLE-	-----	NNNLVYFS	1082	
Db	1049	EKFPAPFQNRAMLVEKTSPT	---	LEVKRKIEIJDGNFYGEDSKINGIPIINDPERSI	1105			
QY	1083	YNNNGALPBMWVPYPAIGSTLGLIAMIITLGAIGIPLAOKRLDCKGKTTFFKAYDTLTA	1142					
Db	1106	FAGFAPALPSWYIPVSVSGVILILLILLGIGIGIPMKRYKVLKLODSSFVDEKFKYDTLTT	1165					
QY	1143	AVGSAYYKKIITQTAWKKKKPALGAGSGDKKPAALAAKPAAPAKRSKASSPAKPTPK	1202					
Db	1166	AVGSAYYKKIITQTSIVIKKAPSAL	-----	KAANNAPKAPVYKPAATJAPRP	-----	1210		
QY	1203	SGAPTKPPAKPPA	1215					
Db	1211	---	PVOF---PKKA	1218				

RESULT 3
T18346
MGC1 protein precursor - Mycoplasma gallisepticum
C:Species: Mycoplasma gallisepticum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18346
R:Keeler Jr., C.L.; Hnatow, L.L.; Whetzel, P.L.; Dohms, J.E.
Infect. Immun. 64, 1541-1547, 1996
A:Title: Cloning and characterization of a putative cytidinesin gene (mgcl) from Mycoplasma
A:Reference number: Z18861; MUID:96201559; PMID:8613358
A:Accession: T18346
A:Status: preliminary; translated from GB/EMBL/DBJ

A.Molecule type: DNA
A.Residues: 1-1122 <KEE>
A.Cross-references: EMBL.U34842; NID:g1022701; PID:g1022703; PIDN:AAB02987.1
C.Genetics:
A.Gene: mgc1
A.Genetic code: SGC3

Query Match	7.8%	Score 501.5;	DB 2,	Length 1122;
Best Local Similarity	22.6%;	Pred. No. 1.1e-18;		
Matches 311; Conservative	174;	Mismatches 470;	Indels 423;	Gaps 68;

```

0Y 5 KKLKRYTLIGLAVGALGASAFGRKQSDKSNUNOVLVQARTILQANSRLAGLQNGSL 64
    ||| : : : : ||| : : : : ||| : : : : ||| : : : :
Db 2 KKL-LFKLSVGTJPLALIGLGSFGLAVSAPKPNKLPMVQGEMLN0-----GOS-WL 52
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 65 FNTVLRDDNNF-----ITANGTILIKLDSFEKPLYG-----LDS 100
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 LEKARWRNNSNTSLSIDCTNGALVLQSKSISRDL-----LYGVIWTFPPGRNNDLT 107
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 101 DDCGY-----KVKQIVSDY-----TTSRNRPDQOTRAYYALVYDEANVHLKRI 146
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 108 GKVGFDYDANNKLTAFSGDVSFNVSJLSSKTVLEATQDQEDPWFVYLLMPDAA----- 160
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 147 NTNSRRIGRNNNSKRFVIGVDNPMAPHVIRFTDGRKFNFTNOTQGEIV-NDFLDAPILP 205
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 -----VQOEQKTKDOVFENYFMSDP--- 181
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 206 KDLPDMYNLYIQRKILPNDVNTAVVPMVGRVSGTNADG-----MFDGDN 252
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 -----AAGDTSAEGSAPFPAGGSGSSAAGGAVAPAPAAASSTARLVEEN 225
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 233 -----GQINTDPIAQT-----KTTDNQNPSTNSGAMPANRHYDQLAV 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 226 SAGMTFMPSTASTSETVIDYNSDONKIPKPKLLDSESSESINGRTYAN--INQKWL 283
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 295 KHRITSFOLDERITMNSNRIGNRN-----NNSKFV--IGVDN-PAHVIRFTDGRKFN- 346
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 QCVI-----VAVNENLEFSENPEFAVENNAFIKPKMVDNYSTPMWQSGANGKMTAV 334
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 347 --FTNOTQGEIVNDFILDAPILPKDL-----HPDMYNLYIQRKILPND 387
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 335 LQFYKHDPNNAVNNRFEYRAKYPRKLETOTPTPLDSSFEYHEPMEY-----E 383
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 388 VNTAVVPMF--VGRVSGTNADGMDHC--GNOQINTPD--PIAQTKTTDNG-----N 434
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 GQDFPMAPMWOYTTLNGGLYAKDGMVYLFEGNGTWTWNNEALSIGVRFKEFENPTAPAGN 443
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 435 PST--FNSGAMPANRHYDQSLNVKHRITSFOLDEKPYVPMTGSEENKNTRLTAGSL 492
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 444 TKTVGTPYGLILSALS-PPATRNGALAPALGQDVGYNF-----VRLAVGV 491
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 493 PSNERWYILDPGTPOYTLKEDSVNVEFSLYLNVSNSLFGDSIYIFGTSE--LPSLM 549
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 492 SSP-----RGANGNIFELGSAI--TWGTNGNGNEFLDTKW 521
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 550 YSFPTRLSDLTALNQVTKDDLEAST--DNGTTNGTNTTADTSFGS-----TGAG 599
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 522 H--SPAVIED-----APPTFTTSSGVQLQNSGSOQSTPMPNNGMESIDPYRTNLSI 573
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 600 TGNNTNTSQVTSNPPLNTYRSFGIDSKPYTSANKIDETNNADNPVI--EARIYAE----- 652
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 574 DYNSTRPALALISKPAAGNFKQ--VESLEFTALKLDLTLN-SLPKKPQENNIFFSYAMLDG 630
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 653 ---RLIGIONE--IPTNAGNFIRNTIGVGFTSTGSRVYLRASTYNG-----DORPFG 699
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 631 ROMSGLKTRDSTWLTNTTINFPYNTQOOLASTAAGENANPRMIINALTTAKGDFDRDRI 690
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 700 NQOPPLY-----VEFGIGYQOQRTGTFW--IGTYLLNSPVDVLDSPRVGTEQNQFR- 750
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 691 NVD-ILISNNTKKFFYY-YQVGCAITTWPEVOVNTKTSANITYYLLTRPDPGSGSTPATQD 748
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y 751 RSLYSPVWGVLTLEGARS--FSNTPYIRAQGDTPESRSIFOSGYSDMTYIEQSVLGF 808
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 749 ANTYSKLNAGVILSTGDOGWYNGSLYKKASFTSSQ-----GYWODEKGLTTTASN 803
 QY 809 DGINNNLVGVKASSPLNSRPNGLMIAATYLRSGIARISGLPNOOPFGTHOV 868
 Db 804 AVISMWTKAGY-----SIRPDD-----TFVSYSKI-----Pf----- 831
 QY 869 ISVSPGDDFSIKIRITTFPGNOJMYFLFTENENKSSVYTLRLADSNPDASSSPSLSL 928
 Db 832 -----EKETIAVANYRSIDS-----YVVOJNSETSVNTVANY-----SPSSA----- 869
 QY 929 IDVNEIGVLLPLDLSFTYNAAGNVALFSSNPGSPTAYVNTFNQINSLDIAFEGSGAK 988
 Db 870 LTLNPKRIINPLMNRD-----NVIGOGAFISRN-DIPSSF-----FENKINDIV----- 912
 QY 989 YTSDFMGCTIOGFDEYLIONGFTSOVARNFVYNOSELSLVDETPA--NAGTNYRVVDP 1046
 Db 913 -----TTEADGTEVIDSKYINSIYRTYPPQNNPHIRLRLVID 950
 QY 1047 DGNLTNQLPLKVOIYLDGKYDAKLNNNLVT--FSTNNNGALPSSWVVPFAIGSTGLIL 1105
 Db 951 RSRATNDFIKILPQV-LVDEYVAVPQANSVFSVSDQETGFDALPGYVLPVAISIPILII 1009
 QY 1106 AIMIILGLAIGIPLRARQKODKGFEKTVDTLTAAGSVYKKIITQTA--NYKKKPA 1163
 Db 1010 ALALALGIGIPISORKKMLKQFALSKKVDILITAVGSVEKIOINTSVTNIKKPPQ 1069
 QY 1164 ALGAGKSGDKKPAAPAKPAPKSPAKPSAPKPTGPKTAP--KPAAPKPT 1220
 Db 1070 MLDGA-----NKKDGA-----SSPSKSPAPAKPPTGPKTAP--SAPGAKPTAPAKKAPAPT 1118

RESULT 4
 A36968
 p1-like adhesin precursor - Mycoplasma p1rum
 C:Species: Mycoplasma p1rum
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-Dec-1999
 C:Accession: A36968
 R:Tham, T.N.; Ferris, S.; Bahraoui, E.; Canarelli, S.; Montanier, L.; Blanchard, A.
 J. Bacteriol. 176, 781-788, 1994
 A:Title: Molecular characterization of the p1-like adhesin gene from Mycoplasma p1rum.
 A:Reference number: A36968; MID:94131957; PMID:8300531
 A:Accession: A36968
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1144 <THA>
 A:Cross-references: GB:LI9685; NID:9404770; PIDN:AAC36866.1; PID:9404771
 A:Genetic code: SGC3

Query Match 5.3%; Score 337; DB 2; Length 1144;
 Best local Similarity 21.5%; Pred. No. 5.5e-10;
 Matches 299; Conservative 168; Mismatches 491; Indels 434; Gaps 74;

QY 5 KKLK---SYLLIGLAVFAGLGSASFGKOSDKSNNTOLVNO-----ARTLDANSVRL 55
 Db 2 KIKRNYKYLILSLISTIVSAALISYSTFNKQJISPIINQNVKSFSPISVGNKVGK 61
 QY 56 AGLGONGSLFNTLVLDVDFNFTANGTIIKLSDFPKYGLD-----LSPD-----CGCY 106
 Db 62 IRHMNNNN--FNGVEIKNGGFVYLSTQATRIDAFGNILWEFDEKIASDSQYANLAKR 120
 QY 107 KVKQIVSDYTSKRNFDQRTAYYALLY-----NDEANVHLKRIINTSNRGNRNNNSK 161
 Db 121 KVEITQDEGDSN-----TLVLLIPKNTPKKQASIDPKLVAYNELTSSKSKQ 172
 QY 162 FVI-----GGVDPNAVIR--FTDGTKE--NFTNQTGELYNDFILDAPLPK 206
 Db 173 ATVOGLIENVNLVYSGSTWPSFTIKLHIDPKKVDVDPNOMKSSQSS-----STFTK 227
 QY 207 DLHPDWNLVYIQRIKLPN--DVNT-----AVYPM-----VGRVSGTNADGMFC 250
 Db 228 EDHPSWVYANNSEKVEHNDQNTNQTNGLKSANMVLPMKQIYTNIGNFAKKGVLITFG- 286

QY 251 GNGQITNDPPIAQT-----KTTDNONPSTFNSGAMPGA-----NNRYSQLVN 294
 Db 287 GNGSLYN--DPEALSIGMKKLDLKPYSQNDNNO-----NKGILPAYLLRLYDPS--- 338
 QY 295 KRIKTSFQDLDRITNSNRGNRNNNSKFEVIGVDNPAHYIFETDDGTKEFNFTQTOGE 354
 Db 339 KPLDGS-----APNR-----RWNQSYAFIGQTDNFTYVPRLANGVOINASTD- 382
 QY 355 IYNDPILDAPILEKDLHPWMLYI-----QRKILPNVNTAVPWPVGRVS 401
 Db 383 -----BAYLYLAAGITVQAKESQAREVISNS--NT-----S 413
 QY 402 CTNADGMFCGNGQITNDPPIAQTKTNDONPSTFNSGAMPGANRYSQLVNKHRIK 461
 Db 414 TNKVYTIQDKRSLQTLGANTITNTKDT-----AANSIDAL-----LFG 453
 QY 462 TSFQDEREYYPWTGSEENKNTRLATGSLPSNERVWIDIPGTPOVTLKEDSVNFSR 521
 Db 454 TAFNIDSLINLPT--TKLENLTIQ--NVQYESYF--DVGATMSVSAVGTYYFBDK 506
 QY 522 -----LYNSVNSLSFIDGSIYIFGTSELPSLWYFPTRLSDTLALNOVKTDIEAST 576
 Db 507 KNHASSTTDINTYTASNGWNMLGRYAF--WSYK--NDIGSIQPKTND----- 555
 QY 577 DNGTTNGTTTADTSSGSGTGNTNTSQTVSNPTLNTY--RSFGIDSKPTSA--NK 632
 Db 556 NNNAI-----YGNLSLILEN--AIWYYSYLSFGSLKCLAGLTK 594
 QY 633 IDETNADPNVLEARVIAEYRLGIONEIPITNAGNFRITIGVGFSTGSRVLRASYN 692
 Db 595 IEMPSKENP-----ENTLYGAM--QVKSILVYLINEBK 625
 QY 693 GDQRPNGNQPPLVYRGVLYGQOTRGTFWYG-----TYKLINSPIYLDSPVGRVETNO 748
 Db 626 SDRSLIAHYHPSISIG-----ESNLVGSAKYQMDYVYKINNNSIGYVSD--YSNTTN 680
 QY 749 FRRTSLYPMWGYLTEEGARSESNF--PYIRAGQDTPESRSIFQSGYSDN-----TYEYI 802
 Db 681 IINGVAI-----YVT--GIKQNDTIPILASQFE-----IGNSPEEDNSSTTKYTGTL 727
 QY 803 QSVL-----GPDGIRNLNVGVKASS--FLNSRPNPGLKIAATYLRSGIARTSGLPN 858
 Db 728 OPTIPMADFQGL--NSTNFENSEISLMLNNQRTKTNNNE----- 764
 QY 859 QOPFGTHQVIVSPGQOFSSIKNIRTFPGNOLM-----YLFETNEN-----KSSYXT 908
 Db 765 -----HFIVTKSP--EISB-----YCGNMIWTEREFYVYNGSNADMGSKRAWFE 808
 QY 909 IRLADSNPDASSFSPTSILDVNEIGVYLLPLDLSNPTVY-----N 949
 Db 809 VK-----DNSLSNSSTTVGWQVIGLDSNLTDADITYVQKNEQPGQGDVILRTRDDT 860
 QY 950 AAGNVALFS--SNPSPG--SYTAVNTFNQNLSDIAFEGSGAKYTSDFMGCTIOFKPDEYLI 1006
 Db 861 OKNNDIFFQJINNTREPGISYCKL--KQN-----YGSFYETISIDPLSL 905
 QY 1007 ONG--FTSOVARNEFTVN-----OSFLNSLVDF--PA 1034
 Db 906 GNGQFVNNLSNOILITNNLNLVQVNLSTVGNPLDSKSTIRYKNOFLNEVQVYKNPT 965
 QY 1035 NAGTNR-----VYVDDGNMLTNQNLPLKVOIYLDGKTYDAKLNN-----NLVTF- 1085
 Db 966 IEGSTPYGVYIVASVNDPVSQSAFTYASWNSLTKYDVMPRTNSSKINVINNSLIFAG 1025
 QY 1086 FGALPSWVVPFAIGSTGLIATMILGLAIGIPLRARQKODKGFEKTVDTLTAAGV 1145
 Db 1026 FSAMADMLPLVLAIPVLAVALITIGGSGIGIPMAKKHAKIKVGELOHDHVGTLTSVG 1085
 QY 1146 SVYKII--NOTAVKKKPAALGAGSGDKKPAAPAKP--AAAKKSPAKKASPAKPTPK 1202
 Db 1086 GVFKKIDTNTSNVKKSPQML-----KAAAKKPTVTPARSQTLNDSVSRPTPS 1136
 QY 1203 SGAPTKPTAPK 1214

Db 1137 S-----APKP 1141

RESULT 5

JUYMAP

adhesin P1 precursor - Mycoplasma pneumoniae

N:Alternate names: cytohesin P1; hypothetical protein 116a

C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C:Date: 31-Mar-1989 #sequence-revision 30-Sep-1991 #text-change 07-Dec-1999

C:Accession: S03725; A38791; J01088; S04516; S62807; S73339; A60102; A27597; P10008

R:Su, C.J.; Tryon, V.V.; Baseman, J.B.

Infect. Immun. 55, 3023-3029, 1987

A:Title: Cloning and sequence analysis of cytohesin P1 gene from Mycoplasma pneumoniae.

A:Reference number: S03725; MUID:88057593; PMID:3119495

A:Accession: S03725

A:Molecule type: DNA

A:Residues: 1-1627 <SU1>

A:Cross-references: EMBL:M18639; NID:g150166; PIDN:AAA25424.1; PID:g150167

A:Experimental source: strain M129-B16

A:Accession: A38791

A:Molecule type: protein

A:Residues: 60-77 <SU2>

A:Note: It is uncertain whether Met-1 is the initiator or whether translation is initiated

R:Inamine, J.M.; Denny, T.P.; Loebel, S.; Schaper, U.; Huang, C.H.; Bott, K.F.; Hu, P.C.

Gene 64, 217-229, 1988

A:Title: Nucleotide sequence of the P1 attachment-protein gene of Mycoplasma pneumoniae

A:Reference number: J01088; MUID:88297153; PMID:2841195

A:Accession: J01088

A:Molecule type: DNA

A:Residues: 1-1627 <INA>

A:Cross-references: GB:M21519; NID:g150138; PIDN:AAA68325.1; PID:g150139

A:Experimental source: strain M-129, ATCC 29342

R:Kenzel, R.; Herrmann, R.

Nucleic Acids Res. 16, 8337-8350, 1988

A:Title: Repetitive DNA sequences in Mycoplasma pneumoniae.

A:Reference number: S04516; MUID:88335593; PMID:3138660

A:Accession: S04516

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1187-1413 <MEN>

A:Cross-references: EMBL:X13087

R:Hilbert, H.; Himmelreich, R.; Plagens, H.; Herrmann, R.

Nucleic Acids Res. 24, 628-639, 1996

A:Title: Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S62797; MUID:96177562; PMID:8604303

A:Accession: S62807

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1156-1199, 'GAT', 1203-1253, 'VRRRS' <HIL>

A:Cross-references: EMBL:U34795; NID:g1215683; PIDN:AAQ43678.1; PID:g1215685

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995

R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73327; MUID:97105885; PMID:8948633

A:Accession: S73329

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1627 <HIM>

A:Cross-references: EMBL:AE000002; GB:U00089; NID:g1673651; PIDN:AA95661.1; PID:g167365

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

R:Jacobs, E.; Gerstenecker, B.; Mader, B.; Huang, C.H.; Hu, P.C.; Halter, R.; Bredt, W.

Infect. Immun. 57, 685-688, 1989

A:Title: Binding sites of attachment-inhibiting monoclonal antibodies and antibodies for

A:Reference number: A60102; MUID:89138634; PMID:2465270

A:Accession: A60102

A:Molecule type: protein

A:Residues: 237-246/702-708 <JAC>

C:Comment: The protein is the major adhesin mediating the attachment of the mycoplasma to

C:Genetics:

A:Gene: P1

A:Genetic code: SGC3
C:Superfamily: adhesin P1
C:Keywords: cell adhesion; membrane protein
F:1-59/Domain: signal sequence #status predicted <Sig>
F:60-1627/Product: adhesin P1 #status experimental <Mat>

Query Match 5.1%; Score 328; DB 1; Length 1627;

Best local Similarity 21.3%; Pred. No. 2, 8e-09;

Matches 268; Conservative 147; Mismatches 404; Indels 442; Gaps 68;

QY	251	GNGQIT--NTDPIA-----QYTTTDDN-----QNPSTFNSGAMPGANN	286
Db	506	GNGHYTKSAHTAPLSIGFVRVRYNATGATVGMPEALLFGSGMVKQTDGLKDLDPFN	565
QY	287	RYSQNLNVHRIKTSQDLERINTNSRIGNRNNKFFYIGG--VDNPAHYRFDPD--	341
Db	566	RW-----FEYVRAVAVAGAKFVGR---ELVLAGITIGDTRVPRLLYDEL	608
QY	342	GTRFNFNTOTGCEIVNDFLL-----DAPY-----LPRDLH--PDVYNL	377
Db	609	ESNLNLVAGQGGLLRDLDLFTPYGMANRPDLPIGAMSSSSSSSHNAPYFHHNPPMDQR	668
QY	378	YIORKITLPPDVNTAVVPMVGRVSGTNADD-----GMFDGNGQITN--TDPIAQ	426
Db	669	PID-----NVYDAFIKPEWE---DKNGKDAKIYIPRYSGMW---AMQVYNNKRLTQD	716
QY	427	KTTTD--NONPSTFNS-----GAMP-----GANN-----RYDSQLNVK---	457
Db	717	PLSADVENAVQPNLSLPAIILNPELLALPDQKVKYKEKEFANERENOKLIVAPPIQ	776
QY	458	-----HRIKTSFO-----LDEKFEVPEMTGSEENKNTIRLATGSLPSNERVWLL	501
Db	777	GTRWMSHFSPILSFGFPNLVSGVLDQVLDVYWMINGRYG-----NNHR-----	822
QY	502	DIPEPTQVTLKEDSVNVEFRLYLNVSNSLST-----GDSITIFGT	542
Db	823	---GVDDITAPQTSACSSSGSISTNTGSGSRFLPTFSGNIGGLKANYQATLGSGQTMITGG	879
QY	543	SELPSLWYSSFPRRLDLALNOVKT-----DDIEASSTDNGTTGTTTATSGSGTS	597
Db	880	S-----PRTILDQANL--OLMTGAGCRNRKSKASSGSDENHTFTSATGMDQGGSG	928
QY	598	AGTGNTTNTSQ---TVSNPTLNTYRSFGIDSKPTFSANKIDETPMAD--PNVIRARIYAEY	652
Db	929	TSAGNNDLSKQDMSISGDSLTQDQNALDQO-----EATVYTNLPRLVPT---ADW	978
QY	653	RIGIONEIPITNAGN-----FIRNTIGVGFTSGSRVYLRASYGDDORPTGCPPELY	706
Db	979	---PALSTFNKNNAQORQLFLRGLGSI-----PVLNRSKSDSNKFOATDQKWSY	1027
QY	707	VFGYLGQQRITGTFWYGTWK--LLN-----NSPYDVLDSPPVGG---TETNG	748
Db	1028	T--DLHSDQTKLNPATVGEVGNLLNPALVETTYGNTRAGSGSNTTSSPGIGKIPDQNN	1085
QY	749	FRRTSLTYP-----VMGCVLTEEGARFSFNTPYRAQGD-----	782
Db	1086	DSKATLITLPELANTPDQVGLVYSGTTVSFQGLGMLV-----FTTDPFKPRAAGVIGLQ	1138
QY	783	-----TPESRSIFQSGYSNDTYEYIOSVLFQDGT-----RNNLVGVKAS	822
Db	1139	LTFGLDASDATORALIMAPRPMAAFRCGSMV--NRLGKREVSVDMLDGLGVADQAOSSQSGSTTY	1197
QY	823	SFLNSNRPNRNGU---EMIAATTYL--RSQIGLARNSGL-----	856
Db	1198	ATRNALPEHPNALAFQVSVVEASAVYKPNSTSGGTQSTNSSPYLHLVKPKVTVQSDKIDDD	1257
QY	857	-----PNQ-----OPEGTHQVIVSPDQSSIKINITIFPGNQLMFLFTNEENKS	904
Db	1258	LKNLLDPNOVTRKLKQSGFTDH---STQPPQ--SLKTTTPVVG-----TSSGNLS	1303
QY	905	SVYTLRLADSSNDPSSFSPTSLIDVNEI---GVILLPLDSSFTYVNAAGVALLFSSN	960
Db	1304	SVLS-----GGAGAGGSSGSGGSGVDLSPEKVSGLVQLPST-----SDGNTS--STN	1351

	Best Local Similarity	20.4%:	Pred. No. 2.4e-08:	Indels	595:	Gaps	82:
	Matches	324:	Conservative	168:	Mismatches	503:	
OY	41	LVNARFLDNLNSVRLAGLQNGSLFNTVLRVYDNNFTTANGTIILKDTKRPYGLDLS	100				
Db	219	LPNGSSSGSDST---NOTGAMFGLKVD---	AVDSSKOSTESLKGESS	263			
OY	101	DDCGGYKKVKQLVDPYTTSTRNRREDQ---	QTRAYYA	132			
Db	264	-----SSITTSSTSTQROGSGSENKRYKALQAVAKKSSQSGNSGDQTEQVE	311				
OY	133	LIVNDEANVHLKRIINTSNKRIKRNNSKFVIGVD---NPAYIRFTDGTCKFNFTN-Q	188				
Db	312	LESDNLANAPIKR-----GSNNQVQLKADFGTAPSSSGSTQGTGTPTPTPL	362				
OY	189	TOGEIYND-----FILDAP-----ILPKDL-----	208				
Db	363	FTTEQIINDPAKFAASILLIYDAPYARRNRTAIDRVHDLDPKAMTAYPPSWTRKMNHGL	422				
OY	209	-----HPDWYN-----LYIQKI-----LPDQNT-----	228				
Db	423	WDMKARDVLLQTTGFNPRIHEMPDGGQYADNEKTEGPDVDSENTNQSGQKADSDKS	482				
OY	229	AVVWMP-----VGRVSGTNADDCMPPCGSGQTT---NTDPIA-----	QRTT	267			
Db	483	APIALPEFAVEANIGLITWEGALLVFG-GNGHTKSAHTAPLISGVFRVRYNATGISAT	541				
OY	268	TDNQNPSTFNFGSGAMPAGNNRYDSQLNVKHKRIKTSFOLDERINTNSRIKRNNSKFVIG	327				
Db	542	VTGWPYALLFGSV---NKOTDLKMLPPNNRNMEIYPPRAAVALGAKFVGR---ELVIA	594				
OY	328	G---VDNPAHYIRFTDD---GKRFNFNTQGEIYVDFIL-----DAPI-----	365				
Db	595	GITTMDDTYVYRLLYDELESNLNLVAGOGGLRDLDTFTPYGMARPLPIGAWSSSS	654				
OY	366	-----LPKDLA---PDWVLYIQKILLPNDVNTFYVWMPVGRVSGTNADP-----	G	408			
Db	655	SSSHNPAPYFHNHPMODPILQ---SVDAFIKPEW---DKNGKDAKTIYPRYKSG	705				
OY	409	MEDCGNGQILTN-IDPIAQTKTTD---NONSTFNS-----GMP-----GANN---	448				
Db	706	MH---ANQVYWNMSNKLTDQPLSADNVNENAYQPSLFAALIMPBLIALLPKRVYKGENE	762				
OY	449	-----RYDSQLNVK-----HRIKTSQ-----LDERFYEPWMTGSBNK	822				
Db	763	FNANEYERENOKLIVAPTOGTWMSHSPSLSFSTGFNLVGSVLDQVLDYVPMIANGRYR	822				
OY	483	NITRLATSLPSNERIWIIDPGTPCVLTKEDSVNVFSRLYLNSVNSLSFI-----	533				
Db	823	G-----NNHR-----GYDDITAPQTSAGSSSGISSTNSGSRSLPFPSSNIGVG	865				
OY	534	-----GDSIYIFGTSLELDSLMATYFPTRLSDTLFALNQVYTD---DIASSTDNG	579				
Db	866	LKANVAQNLIGSQOTMITGGS-----PRRTLDQANL-QLMTGAGMWRNDKASQSGD	914				
OY	580	TTTNGTGTADTSSSGSTGAGTGN-----TTNTSQYVSNPFLNTYRSFGI	623				
Db	915	DHTEFISATGMOGOEDSGTSAGNPDSLKODKISKSGDSLTDGDGAMMQOATNYTLRP	974				
OY	624	DSKPT-----SANKIDENFMW-----	638				
Db	975	NITFLADMPALSFETKNNNAQRAOLFGLGLSGIPVLVYKSGDDNSKFKADOKMSTYD	1034				
OY	639	---ADPNVTEARIYAEYRLQIONEIPI-----TANGNFINTIG---GVGF-----	TST	681			
Db	1035	LOSDQTKLNPAYGEVN-GLLNPALVEYFGNTRASSGSGNTSSGIGIFKIPEDSGTNT	1093				
OY	682	GSRYVL-RASINGDORPTGN-----FO---PPLYE-----GYGVGO-----	714				
Db	1094	TSKAVYLITPEGLAMTPODYGNILVYSGSFSFOGLMVLFTFTFKIPRAGYLGILQTLGDAS	1152				
OY	715	QRTGTFTVYGTLYKLNNSPYDLDSPRV-----GTETNOERKTSILYPMWGVGLTEBBA	768				

```

Db 1154 DATA:ALIM-----APRPAARAGSGVMNRLGHEVSWDLKGVWADQ--A 1195
OY 769 RSESNPTPIRAGDT-PESSIFQSGVSDMTVEYIOSVGLFDFIRNNLNVG---VKASS 823
Db 1196 QSSQSSQSTTATRDALPE-----HPNALAFQVSVVSAAYKPPSTSGGTQSTNSSP 1246
OY 824 FLNSNBNP-----PNCLEMIATATYLRSGTGLARGLPNOQPFCTTHQVTSVSPGD 875
Db 1247 YLHLVAPKPKVIGSDKLDLDDKLNLDPNQVTRKL-----RQSGGDTH---SQGPQP 1293
OY 876 QESSTINRTIPFGNOLMTFLFTNNKSSVYTLRLADSNPDASSFSFTSLIDVNEI- 934
Db 1294 Q--SLKTTTPVFC-----TSSGNLSSVLS-----GGGAGGSGSGSGSGVDLSPV 1337
OY 935 ---GVILPLLDNSFTYVNAAGNALFSSNPGSGSYAVTQNG-----NLSDIAPEGSG 986
Db 1338 KVSGLVGLQLPST-----SGGNTS--STNNLAPNT-----NTGDVYGVGLS-----ESNA 1382
OY 987 AKYTSDFWGTIQKPEYLLIONGFTSQVARNFVT-----NOSFLNSLVD-ETPAN- 1035
Db 1383 AKMNDVDGIVRTPLEALLDGEQADTGQSVKFKSPQIDFNRLFTHVTDLPDVTM 1442
OY 1036 -----AGTNRYV-----VYDDGNTLNTQNLPLKVOIQIYLDGKYYAKLKN- 1076
Db 1443 LVYDQYIPLFIDIPASVNPMMVRLKYLSEFDTN--EOSTGLRLEFFKPD--ODTOPNNV 1497
OY 1077 -----NLVTFSS-----YNNFGALPSWVVPALIGSTGLAIMIILGLAIGPL 1119
Db 1498 QVNPNNGDPLPLTASSQSGQRTLFSPNOMPDYVLPALIVPVIYVLSLTGLAIGIPM 1357
OY 1120 RAQRKLQDGFKFTTKKVDLTLTAAVGSYVKITITOTANVKKRPAALGAGSGDKKPA 1179
Db 1558 HKRQKALKGAFALSNQKVDVLTAKAVSVEKEIINRT-GISQAPKRL-----KQTSAA 1608
OY 1180 KPAAPAKPSAPKASSPAKPTGPKSGAPTKP 1209
Db 1609 KPGAPRPVPPKPGAPKPPVQP-----PKKP 1634

RESULT 7
A30588
140K adhesin precursor - Mycoplasma genitalium
N:Alternate names: attachment protein MgpA
C:Species: Mycoplasma genitalium
C>Date: 08-Jun-1989 #sequence, revision 01-Dec-1995 #text_change 20-Apr-2001
C:Accession: A30588; J00090; B64221; A61605; J18721
R:Ballio, S.F.; Chavoya, A.; Su, C.J.; Baseman, J.B.
IntAct: Immun. 57, 1059-1065, 1989
A:Title: DNA and protein sequence homologies between the adhesins of Mycoplasma genitali
A:Reference number: A30588; MID:89173298; PMID:2925238
A:Accession: A30588
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1444 <DAL>
R:Imamine, J.M.; Loeschel, S.; Collier, A.M.; Barille, M.F.; Hu, P.C.
Gene 82, 259-267, 1989
A:Title: Nucleotide sequence of the MgpA (mnp) operon of Mycoplasma genitalium and comp
A:Reference number: J00090; MID:90060815; PMID:2563522
A:Accession: J00090
A:Molecule type: DNA
A:Residues: 1-1444 <INA>
A:Cross-references: GB:M31431; NID:g150157; PID:AAA25420.1; PID:g150159
A:Experimental source: strain G-37 (ATCC 33530)
R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhmann, J.; Nguyen, D.; Uterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.;
C.A.; Venter, J.C.
Science 270, 397-403, 1995
A:Title: The minimal gene complement of Mycoplasma genitalium.
A:Reference number: A64200; MID:96026346; PMID:7569993
A:Accession: B64221
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1444 <TIGR>

```

```

A:Cross-references: GB:U39696; GB:I43967; NID:g1045869; PID:g1045876; TIGR:MG191
A:Experimental source: strain G-37
R:Mader, B.; Hu, P.C.; Huang, C.H.; Schillz, E.; Jacobs, E.
Zentralbl. Bakteriol. 274, 507-513, 1991
A:Title: The mature MgpA-adhesin of mycoplasma genitalium.
A:Reference number: A61605; MID:91321682; PMID:1863319
A:Accession: A61605
A:Molecule type: protein
A:Residues: 59-68 <MAD>
R:Peterson, S.N.; Schramm, N.; Hu, P.; Bott, K.F.; Hutchison III, C.A.
Nucleic Acids Res. 19, 6027-6031, 1991
A:Title: A random sequencing approach for placing markers on the physical map of Myc
A:Reference number: S18693; MID:92051396; PMID:1945886
A:Accession: S18721
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 106-176 <PER>
A:Cross-references: EMBL:X61522
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, September 19
C:Keywords: membrane protein
F:59-1444/Product: 140K adhesin #status experimental <MAT>

Query Match 4.48; Score 285; DB 1; Length 1444;
Best Local Similarity 20.98; Pred. No. 4.4e-07;
Matches 270; Conservative 159; Mismatches 477; Indels 386; Gaps 68;

24 SASGKFGKQSD-----KSNMQVLVNAQRITLDANSVRLAGQNGSLNFTLRVDND 75
Db 418 NTSPEGVGDTDHKKDGFKNSSPIALPREAYFANIGNVAIG-----NSV----- 464
OY 76 FITANGTILKIDSTFKPLGLDSDCGGKQVLYSDYTSRRNRFQROTAY-YAL- 133
Db 465 FIFGNGHAKMT-TNPL-----SIGVFRK-----YT-----DNRSKSSGVMPYAVL 508
OY 134 ---LVNDEANVHLKRITNSNR-----IGNRNNSKFVIG--VDNPAH 172
Db 509 FGLINPQTN-GKLDPLGTRNWFYVPRMAVSGVKVGN-----QLVLAGTILMGDTAT 562
OY 173 VIRETDDGTR--FNTNQTOGEIYNDFILAPILPKLHWDWNLTYQRKILPNDVNTAV 230
Db 563 VPRKLYDQLEKHNLVAQGGGLREDLQITPY-----GMAN---RPDIP-----V 605
OY 231 VPPVGVSGTNNDDGFMDCGNGQITWTDPIDIAQTKTTDQNPSTFNSGAMPAN---NR 287
Db 606 GAW-----LQDER-----GSKFGPHYFLN 625
OY 288 YDSQLNVKRIKTSFOLDERINTNSRIGRNNNSKFGVGNPAHVIRFTDQ---T 343
Db 626 PDIODNNNN-----DIVEALISSYKTKDKL-----KHVYRYSGLYAWQ 665
OY 344 KKFNTNO-TOGEIYNDFILAPILPKLHWDWNLTYQRKILPNDVNTAVPVPVGVSG 402
Db 666 LFNWSNKLNTPTLSPANVNSYAPNSLF-----AAILNEDLTGL----- 706
OY 403 TNADDGFMDCGNGQITWTDPIDIAQTKTTDQNPSTFNSGAMPANNRYSQNLVYKRIK 462
Db 707 ---SDKFFYKKEHFAENEDRFQQLSLNPN-----NTMAAYLVANVQKFTT 752
OY 463 SFOLD-----EKVY-YEPWTSSEENKNTIRLATGSLPSNERKWIIDIGTPQVTLKEDSVN 517
Db 753 GPNLDSSTFQDFLDFLWION--GKPSNSPSPSTASAS-----STPLPFPFSNINVG 802
OY 518 VFSRL--YLVNSVSLSTIGSIVYFGSELPSLV---YSPFTRLSLDLALNOVKTD-- 570
Db 803 VKSMITQHLKKEKTR-----WFLPNESPDIWTGAGYVQASNOKNGIPEQVYKPSN 855
OY 571 ---IEASSTDNGTNGTNTTADT-----SSGSTAGTGTNTNTSQTAVSNPTLN--TYR 619

```



```
QY 1019 VTNQSLNVLVDFTPANAGTNRVVDPDGNLTNOMLPLKVOIQYLDGKYYDAKLNNML 1078
DB 902 TQNA-----SAVFETNPNVVTG---AIDNTGN-----ANNNG 930
QY 1079 VTESYNNFALPSWVYPTAIGSTLGLAMIIILGLAI--GIPLR--QRKLOKGEFTTK 1135
DB 931 VFTFGS-----TYTGNIGNTNALATISVGAGKATGCAIKATTKTLTDMAQSVFT 983
QY 1136 KVDLTAAV---GSYKRIITOTAN 1157
DB 984 NPVYVTCALDNTGNANNGIVTFTGD 1008

RESULT 9
AG2560
hypothetical protein al18078 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120g4
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2560
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriuchihara, S.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2560
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 11487 <KUR>
A:Cross-references: GB:AP003603; PIDN:BAH77408.1; PID:g17134851; GSPDB:GNO0182
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al18078
A:Genome: plasmid

Query Match 3.8%; Score 245.5; DB 2; Length 1487;
Best Local Similarity 21.0%; Pred. No. 5.7e-05;
Matches 233; Conservative 163; Mismatches 454; Indels 335; Gaps 64;

QY 30 KOSDKNDMTOLVNOART--LDANSY-FLAGLGNGSLNFTVLRDVEDNPTTANCTTIK 86
DB 453 KISNPGEIGIYLENIGTITHSNTNNVGNAINLSQVST--ETTNLTNNITGTGIL 510
QY 87 LD--SPTKPLXGLSDDCG-GYKVAQIVS-----DYTTSRNR--FDQKOTRAY----Y 131
DB 511 LFNVPDSTSTNTNTNTGTDGIGLSQVISPITNNSTNSGRGFIENSTGTINSDN 570
QY 132 ALLVDEANVHLKRT-----NTNSNRIGNRNNNSKEFYIGVDNPAHYIRFTDGTKEFTN 187
DB 571 TLTDDNQEAIKLVEVGTITINQNTINGSNNSF---GID-----IANTGDIINLTIAN 620
QY 188 QTQGEIYNDFILDAPILPKDLHPDWNLVYIQKILPNDVNTAVVPMVPGVSGTNNADGM 247
DB 621 NMLSEFNPD-----IRVDLSTTGTCTLOINNNTIISNH-GTALDLOLVENTNLTA---- 668
QY 248 FDCGNGQITNTPIAQTKTTTNDQNPSTNSGAMPANNRYDSQLVWKRIKTSFOLDER 307
DB 669 -TINNNOITGT-----ATSTTSQGIHLQAVDMAIIGFVEVDNTISMNTGDDGNGFQNGS 723
QY 308 INNSNRIGNRNNNSKFVIGGVNPAHYIRFTDGTKEFTNTOGEIYNDFILDAPILP 367
DB 724 STATFIAGNTISN-----IOTDTPG--DFNTEGIVNEVEFNAAKEITVAN 768
QY 368 KDLH--PDWYNLYIQKILPNDVNTAVVPMVPGVSG--GNMADGMEF--CGSQITLND 421
DB 769 NNTSQVPVNGIIMQAPAGDNAGSIDITDNTISQITTEGNGIDLQFFDNFDSNTNNTTV 828
QY 422 PIAQT-----RTTNDN-----QNPSTFNSG--AMGANNRYDSQLNVKH 458
DB 829 NTAARNISDIQGTGINAFNETPNTIIVYDQNTISSGDDGTLNGVNNAAQVYINNNT 888
QY 459 RITSQDLDEKFEVPEMTGSEENKNTRLATGSLPSN-----ERYWIDIGTPOVYL 511
```

```
DB 889 AGEQGIQV-----ENAN-----GTINVDNSTITTDQOESISLSEVGT--VSU 929
QY 512 KEDSVAVFRLYLNVSNSLSFIDSDIYIFGTSELPLMWYSFPTRLSDLTALNQVKTDDI 571
DB 930 SONIVN--GNNHNGITITNIGQESMLNNGNTLTEN--ENDIRVSLNSTTGYEILNNIT 986
QY 572 EASST-----DNGTTTNGT-----TTTADTSSGSF-----GAGTNTTISQTVS 611
DB 987 ENSGTALIDLQEGNTMLTATINNNOITGTATSTSGIINLQADMAAIGDFEVSQNTIS 1046
QY 612 NPTLNTYRSFGIDSKPTSA-----NKIDENMADPNVIAEKRTIAEIRLGIQNEIPT 663
DB 1047 NMTNGDGTNQLGSSSTAFETNAGNTISMIDTDPDPFNET-----GINVEV--- 1094
QY 664 MGNFIRNTIGVGFSTGSRVYLRSYNGDORPTEGDFELVFGYLG-YOOTRTGTFM 722
DB 1095 -----FNAAKEITV--ANNISQVPVN-----GIYMOFAGD-- 1126
QY 723 YGTYKLLNSPYDLDS--PRVGETN---QP-----RRTSLTPYMGVGLTE--E 766
DB 1127 -----NAGSIDITDNTISQITTEGNGIDLQFFDNFDSNTNNTVNIARNISDIQGT 1179
QY 767 GARSESNTPIYIAQGTPEPSRSTFGSGYSDNTYETIQSVLGFPGCIRNNLVGKASSFLN 826
DB 1180 GINAFNKT-----DNTIIVVDQNTIS--NSFR-----DGI--NLN-GVNNAAQVIN 1221
QY 827 SN-----RPNNGLEMIATYILNSQ---IGLARTSGL-----PNOQPGTTHOYV 869
DB 1222 NNTTNAGEGIGVENVNGTINVDNSTITTDQOESISLSEYTGVSQNTANGNNHNGI 1281
QY 870 SVSPGDQFSIKNIRTIFFGNQLMWELFTNENN-----KSSVYTLRLADSSNP----- 918
DB 1282 AIA-----NTSSTNLTLY--NSNQ-----LTENFNDIRVLSPTIGTILQINDNTISNCTAMD 1332
QY 919 ----ASSFSPTSL-----IDVNEIGVILLPLDENSEFYVNAAGNVALBSNGSPG 965
DB 1333 VQCGANTNLVTSILANNQITGVADITSSGILNQAFDNV-----VAGNVT----- 1377
QY 966 SYTAVNTPFQNLSDIAFEGSGAKYTSDFMGTIOFKFDEVYLIQNGFSGVARNFVFNOSFL 1025
DB 1378 --AAGTIN-----TTNGDGMNFQNGSTAF-----QATNTIISN----- 1413
QY 1026 NSLVDETPANA-----GTNVRV-----VYDPDGNLTNOMLPLKVOIQYLDGKYYDAKLKN 1075
DB 1414 --IROTADAGANFTDGINFIDPAAGSYELSNVVDIQGYGISITIFSSGSVNTIISN 1471
QY 1076 NMLVT 1080
DB 1472 DMOIT 1476

RESULT 10
D71917
toxin-like outer membrane protein jhp0556 - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: D71917
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
A:Reference number: A71800; MUID:99120557; PMID:99233682
A:Accession: D71917
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3194 <ARN>
A:Cross-references: GB:AE001488; GB:AE001439; NID:g4155100; PIDN:AAD06134.1; PID:g4155
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp0556

Query Match 3.8%; Score 242; DB 2; Length 3194;
```

Db 901 DAFNLKNNNNNSNTVFYLGNM-----TTWKTACTGVFTQDYGSTNSVLVENQTTPFLAG 955

QY 859 QQPEGTTHOVISY--SPGDQFSIKNTIRTPGNOL----- 892

Db 956 ANP--TNSVGFGRKISGAEWGLVGYIGCVKAKNOIDITGRISGNGAKTGGATLVNA 1013

QY 893 -----WYLFETDENKSSV-----YTRL----- 911

Db 1014 QERLNTANLNLNDRKAGLQNSWMNFIYVNGMLNTNANFSQTPHGGFNLKANNTWDKG. 1073

QY 912 -----ADSNPPASSFPTSLID-----VNEIGVILPLDNSF--YTYNAA 951

Db 1074 SVSGCGNFVGDVANNANNAVIRKVNFSDDGLITLYGGENSAGNSLTLENNFNSYNMAK 1133

QY 952 GNVALFSSNPGSPGSY-----TAVNT-----FN--ONLS 978

Db 1134 AONLIIFNNNSFGSYSFNDKTKVTEKGTNTLTNSDPSRLKGSVIDNNSIFNIERDLT 1193

QY 979 D-----IAEGSGAKY-----TSDFWGTIOFKPDE-----YLQNGFTFSOV 1014

Db 1194 DKTYYTLTSGNITKYNNOALADNVFSKNLMDLIHYDEGTLRLTDNNITFVO--FTQSN 1251

QY 1015 ARNETVNSQFLNSLVDETPANAGTNRVY-----VDPGG--NLTONQLPLKVOIYLDGK 1067

Db 1252 GQKFVEEET-----FPGSITRYKYPTTISSPPTHEADSKDITMNG--VRKQPFPIGK 1301

QY 1068 -----YDAKIKNNNLY--TFSYN--NGFALPSWVVPATISGLITLAIMIIGLAI 1116

Db 1302 TPVCYGVCCITAYAKNODLIGSAPFAMSLNFGA-----TVVGT-----LLGS----- 1342

QY 1117 IPLRAQRKLQDKGFKTTFKAYDTL 1140

Db 1343 ---SAQERKANNGSGIMFGKNLTL 1363

RESULT 11

576412
hypothetical protein slr0408 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_Change 20-Jun-2000

C:Accession: S76412
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-116, 1996
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: *Synechocystis* hypothetical protein slr0408

s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1199 <RKN>
A:Cross-References: EMBL:U90915; GB:AB001339; NID:91653604; PIDN:BA18541.1; PID:91653604
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: *Synechocystis* hypothetical protein slr0408

Query Match 3.6%; Score 233.5; DB 2; Length 4199;
Best Local Similarity 20.6%; Pred. No. 0.0011;
Matches 264; Conservative 161; Mismatches 525; Indels 409; Gaps 66;

Db 16 LAVFGALGSASRG-----FKQSPKSDNTOL-----VNGARTLLANSVR 54

QY 531 VSLIGLIGRIAGFPPTWTQTFPAASSQNNQGLEANNVTASVDFGLTLINLAQLEPN--- 587

Db 55 LAGLGGNSGLFTVLADVDNETTANGCTITKLDSPFKPLXGLDSDDCGKYKQIVS- 113

QY 588 ---LSLNPDEFTVT-----VTNANQI-----STLPGVGVVQDSDSSNNSEVILRL 631

Db 114 ---DYTSANRPDQQTIRAYVAILVND--ANVHLKRIKNTSNRIGNRNNSKFTYIGV 167

QY 632 EOEIPYTT-----VNGQPFENIIIV--YSGNQLGN-FTSSPVALQSP 670

Db 168 DNPAHVIRFTDQGTKEFNT--NOTQGEIYVNDITLADILPKDLHPDWNLVYIQRIPLPDV 226

576412 *Synechocystis* sp. (strain PCC 6803)

C; Species: Synechocystis sp

A; Variety: PCC 6803
Date: 25-Apr-1997 #sequence

C;Date: 25-Apr-1997 #sequence_2000000000
C;Accession: S76412 E; Nakamura, Y.;
Asamizuo,

C; accession. Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, T.; Kaneko, T.; Sato, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Naka, T. 1996. *Journal of Eukaryotic Microbiology* 43: 109-136.

A: Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechococcus* DNA Res. 3, 109-130, 1990

S. 674332. MIMD:97061201; PMID:8905231

A;Reference number: S74322; MUID:9/001201, INDIANAPOLIS
Accession: S76412

A:Accession: 5/0412
A:Status: preliminary

A: Molecule type: DNA
PDB: 9165

A;Residues: 1-4199 <KAN>
A;Cross-references: EMBL:D90915; GB:AB001339; NID:g1653604; PIDN:BAAL8541.L; PIV:g1653604
A;Cross-references: EMBL Data Library, June 1996

A; Note: the nucleotide sequence was submitted to the EMBL Data Library, accession number X61408

C; Superfamily: Synechocystis hypothetical protein s1000

Query Match 3.6%; Score 233.5; DB 2; Length 4199;

Query match	20.68;	Pred. No. 0.0011;
Best Local Similarity	161;	Mismatches 525;
		Indels 409;
		Gaps 66;

Matches	284;	Conservative	161;	Mismatches	227;
-----VNOARTLDANSVR 54					

16 LAVFALGSASF-----FKQSDKSDNTQL-----VNUAKRLPDANSVN 24

[illegible]

Db 531 VSLGLIGARIAGFPIWVQIFAASSQDNGV... 113

55 LAGLGQNGSLFNTVLKRDVDNFIITANGTIKLDSEFKPLXGLDLSDDCGIAVAVQAVS

[illegible]

Db 588 ---LSLNPTDFIVI-----
 LKRNQNSNPIGNRRNNNSKFVIGCV 167

114 ---DYTTSRNFQRTRAYYALLVNDE--ANVHLKRINTNSNRIGKNKNNNSAEV L009 100

670
 FTSSPVALQSP
 YSGNQLGN
 VNGQPFNSNIIVK

 FQETDVT
 670

[illegible]

168 DNPAAVIRFTDDGTKENFT-NQTOGEIVNDFILDAPILPKDLHPDWINDI


```
Db      671 DLILYINPT-SGTNDYQNNQOITLAFNGCLDPLDILPTE-----NOLLMWFATVNSS 723
QY      227 NTAAYPMWPGRGVSGINADGMDGCGNGOITNDPIAOTTTDNDONP-STFSGAMPKAN 285
Db      724 NQALITDVA-----VNOSRVLVLTASMTAIPGEIYVQYNPDAN 764
QY      286 NRYDSQLNKHRIKTSFOLDERINTNSNRGNRNNNSKVFVGVDNPAHVRFTDGTGF 345
Db      765 S--DQOLKRAQ-KITLAFVTSNGTAPMTAGLINTFAGVINVOQPV-LSNLITSD----- 816
QY      346 NPTNOT-----QGEIYNDFLDA-PILPKDLHPWNLXIQRIKLPNDVNTAVVPM 396
Db      817 -FAODTSPALALTSOGDILLAMSSDTPPTIPISVLAEGDYLYL--VFADNLKDSANPSS 873
QY      397 VGRVSGTNDGDMFCGNGOITNDPIAO---TKTTDQNSTFSGAMPKANNDYQSO 453
Db      874 NSQFTIKTSD-----GNTTTPPNVSLAONTITLTLTNSQIVVSVLSGTNLITN 927
QY      454 LNVKHKIKTSFOLDERFVPEWTSSEENKNTRLATGSL-----PSNERYWILD 502
Db      928 LVLADATNNSFVW-----PDFTNSVQASGSTAPSTLLGSVISLITLIPFNQITLVNQ 981
QY      503 IP--GTPVYTLKEDSVNVSRLYLSVNSLSFIGDSI-----YIQTSELPSLWYSFPT 555
Db      982 IFNGGQFYVTANGNSTSPI-----TVTSVTAVADTSITLVNLQIIGOGQVLTITY--TPN 1033
QY      556 RLSDDLALNQVKTDOLEAS-STDNGTNGTITTA-----DISSSTCA 598
Db      1034 SGENNNLVNSTGTPQTVASFSTNNELTASSTGTGYKTAISPFGDGSISITTIPTGTI 1093
QY      599 GTGNTTNTSQTIVSNPTL-NTYRSFGIDSKPTANKIDETNMADPN-----YI 644
Db      1094 NSDVVALTYLQNSTKALQNVVAMVNDTSALSUKITPGQNYG-PNPAALITTAQOSDLY 1152
QY      645 EARIATVRLGIGONEPIITNAGNFRINTIG-GVGFSTGSRVYVLAASVGGDQRPNGRNP 703
Db      1153 YAVLPGDNQWGLAAPFSSQPOGQDKVTLGVPBGMLLAWMLTOLDSDGPDPTTQIOLAT 1212
QY      704 FLYVEGYLGQOTRTGFYGYTKLLNNSPYDLDSFRVTEGTNOFRSTL----- 754
Db      1213 F-----KOTTM-----TMP-TILGANGANGINPNSSELSISSINGOPAI 1250
QY      755 -----TPYMGGYLFEGAR 769
Db      1251 FWTESRPSYSLVSEONPLVYLRGLSELGTIVINQQLSVANGTYSTAGTYLGGVGL 1310
QY      770 SFSNT-----PYIAGG-----DIPESRIFQSGYSDMTYEXIYGLGPDGIRN----- 813
Db      1311 EMTNTNTGDFNPVAVLFGGGGITTINSFVPSV--QGFSEVFWKLPSTDGVGLANLAGVF 1368
QY      814 NLNVGKASSFLNSNRPND-----NGLEMLA 839
Db      1369 DLSLNDLSLITLNNNSNPQISGYITTTGMMHYVVCYDVKOILDXLDQVLNTLENTIA 1428
QY      840 ATTYLRS-QIGLARTSGLPNOQPFQTHVOIS-VSPGDQFSIKIRITLFFNGQL-----W 893
Db      1429 FANLPQSGTLTLGSGSVYLDDEFAYNSILSYDNGSSPSSNNFNLMTGSLNGIWM 1488
QY      894 YLLFTNENKSSVYTLRLAD--SSNPDASSSF-----SPTSLIDVNE-IGVYLLPL 941
Db      1489 -----GVNEVGSHYQARFEPEYTAGPETNYSWDDSSGNSWQSVSINPDEVVPTLLSAA 1543
QY      942 DNSFYTV--NAAGNALFSSNPGSGSYTAVN-----TFNQLMSIDIAE 983
Db      1544 NNPIMDIVSANPAGANNNAQIAINGNPDIIFQVNLITGOQSEITGFTVTTSSNNQMLMVGD 1603
QY      984 GSGAKYTSDFW--GTIQKPEDEYLLIONGFTSOVARNFVTNOSFLNSL-----VDTPPANA 1036
Db      1604 GTGNMVF-SESQOLGYI-----LAENADSTTPOLEFISGDKLNLNGCATFSPSHRVMA 1655
QY      1037 GNNRYVVDPDGN-LTNQNLPLKVOIQYLDGKYDDAKLNNNLVTTFSYNNFGALPSPWVP 1095
```

```
Db      1656 TETFTIYVDEGSPLTS---PATVNI-YLQGO-----TDPITFT--SLSPINOGGP 1701
QY      1096 TAIGS-----TLGLAIMILGLAIGPLAORLQDQKGRTEFKVDTLTAAGVSV 1147
Db      1702 VASNPDPYLDNOVYLGIATIKEAND-----ASLSLVDSGF-----VIDTNPALIAAV 1747

RESULT 12
A:97859
199-Kha cell surface antigen [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence-revision 30-Sep-2001 #text-change 30-Sep-2001
C:Accession: A97859
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fourminter, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97859
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <KUR>
A:Cross-references: GB:AE006914; PIDN:AA03811.1; PID:g15620410; GSPDB:GN00173
C:Genetics:
A:Gene: rompa

Query Match
Best Local Similarity 21.0%; Score 232.5; DB 2; Length 2021;
Matches 278; Conservative 147; Mismatches 536; Indels 365; Gaps 64;

QY      9 SYTLIGGLAVFALGSASFQKSDKSNQNTQVNOQARTLANSVRLAGLQNGSLNTV 68
Db      162 NYTGGLNIALGA-----NAALIIQSA--PAKITLGA-NINGGCIITV 202
QY      69 LRVDVDFITANGTITIKLSFTKPLYLGLDSDGGRVVKQIVSDYTSRRRFQGR 128
Db      203 KTD-----AATNGTGTNTALATVNVGAGIA-TLEGALIKATTTKLTNAASVLTITNVN 255
QY      129 AYVALLVNDKANV-HLKRINTN-----SNRIGNRN-----NSKFIYGVDPNPAHY 174
Db      256 AVLTGAIDWTGVDNVGVNLNGLALSQVGNIGNTNALATISVAGAKATLGGAVIKATTT 315
QY      175 RFTDGTGRNFNTNQ--TQGEIYNDFLDAPILPKDLHPWNLXIQRIKLPNDVNTAVV 232
Db      316 KLTDNASAVTFNPVNVYTGAIIDN-----TGANNGIYV 348
QY      233 WP-----VGRVSGTNDGDMFCGNGOITNDPIAOTKTT--TDNONPSTF-NSGAMPGA 284
Db      349 FTGDSIVYTNIGNITNA-LATISVAGAKATLGGALIKATTTKLTDNASAVTFNPVNVYGA 407
QY      285 NNRYSQNLVKKHRIKTSFOLDERINTNSNRIGNRN-----NSKFIYGVDPNPAHY 337
Db      408 ---IDNTGNANGIYV-FTGDSIVTGN---IGNTNALATISVAGAKATLGGALIKATTTK 460
QY      338 FTDDGKREFTNQ--TQGEIYNDFLDAPILPKDLHPWNLXIQRIKLPNDVNTAVV 395
Db      461 LTDNASAVTFNPVNVYTGAIIDN-----TGANNGIYV 493
QY      396 P-----VGRVSGTNDGDMFCGNGOITNDPIAOTKTT--TDNONPSTF-NSGAMPKAN 447
Db      494 TGDSTYTGINGINTNA-LATISVAGAKATLGGALIKATTTKLTDNASAVTFNPVNVYGA- 551
QY      448 NRYDSQLNKHRIKTSFOLDERFVPEWTSSEENKNTRLANGSLPSNERYWILDPGP 507
Db      552 --IDNTGNANGIYV-FTGDSIVT--TGNGINTN--ALATISV-----GAG 589
QY      508 QVTLKEDSVNVSRLYLSVNSLSFIGDSIYIFQTSLSLWYSPPTLSDLPALNOVK 567
Db      590 KATLGAIIKATTTKLTDNASAVTFN-NPVNVYTGAIIDN-----NTGN 629
QY      568 TDDIEASSTDNGTNGTGTTTADTSSTGAG-----TGNTNTSQTIVSNPTL 615
Db      630 ANNGIYTFIGNSVIGNIGNITNATLVNVGAGIATLEGAVIKATTTKLTNAASVLTITNV 689
```

[illegible]

QY	2	NISKRLK---	SYTLIGGLAVFALGASAFGEFQOSKSDNDTQUNQAKRFTLDANDANSVRLAGL	58
Db	2027	HISKRLKFKLEQNIENVEVYSPILTN-	-----SIVNNKNVNTN--FTLD-----	2070
QY	59	GOMGSLFNFLRVDNFNTIANGFTIKLDSFTFKDYLGLDSDCGYKQVSDYTT	118	
Db	2071	---NLKSNQLYRLVDVYIIDDNNDTINDKNKPKANNTTRILDIAPG---	KTTS---KS 2121	
QY	119	RNRPDQROTBRAYALLVNDDEANYHLKRINTSNRIGNRNNNSKPYIGVDNPAHVIREFTD	178	
Db	2122	NNTWNTSTSSOEFEVINSD-----	DENEY-----LDNLKATISFK 2158	
QY	179	DGT-----KRNFNQOGELVNDFFILADPLFKDHLHPWYNYLQIRKL---	PNDVNTAV 230	
Db	2159	GOTLLTPEVKVNIILKQN-----NKYLLKQI--TNLEPE--	NRVYVESILLAKPKTKKPL 2209	
QY	231	VPMFVGVHSGTNNDDGDFDCGNGQITNPDIAOTKTTDNOQSPFNSG-----	AMPGAN 285	
Db	2210	V-----VELINKDISFOTOAGNY-----KVIOIK---	SQNPSTVDYKORIKLKLDDIQ 2255	
QY	286	NRV--DSQLANV-----	KHKRITSFOLD-----ERINTNSNR 314	
Db	2256	NAMEKOLEITYSANDNSTYTAIIKLEKNLEFEFELTNLEKNRITYTFEKIEDINDNTK	2315	
QY	315	IGNRRNNS---KFYI-----	GVYDNPAHV-----IRFDDGKRFNFTQOGE 354	
Db	2316	TPFKNSDSIDQKFLVLSNNOGVGNIETIODRVDNHLNSAKIRFELINDLVLSDEQAT	2375	
QY	355	IV--NDFILDPIL-----	PKDLHDMVNLTKORILPNDVNTAVPMFVGVHSGTNNADD- 407	
Db	2376	IYNNNOFTTSKAVITDQNOKYLEATFNSLVNLKNDTIINKLEFNTKRNKSNKINGINDTV	2435	
QY	408	-----	GMPCGNG 415	
Db	2436	IYDATNLIINDKLTGPFLTKFEPEANKNTNISVLELDTNNHISKNTLYTAKPFNSDG	2495	
QY	416	QITNTDPLAOKTTTDNOQPS-----	TENGAMPG-----ANRYD 451	
Db	2496	OSVLNTPDIADKITYNNNNKKELTNLTNLTSKROIIFKGIYVNSQNTIDENKFE	2555	
QY	452	SOLNVYHRIK-----	TSFOLDEKFYVPEMT-----GSEBKNIT- 485	
Db	2556	KNSNDVUKITYKPTITTIQKNGNMTFOPQNAQKRFNINSNDVDFSTDIADITFESDOH	2615	
QY	486	---RLATGSLPSNERY-----	IIDIPOTQVTLKEDSV-----NVFRELINSVNSLSFT 533	
Db	2616	DTKKTITTLKOKDNQWYIEDTISLAYNDYTKLDDISITPKMAFANLKIQLEN-----	2670	
QY	534	GDSIYIEFGRSELPSLWYVSFTYRLSDLALNOVKTDIEASSTDNGTTNGITTTADTSS	593	
Db	2671	-----	KEQISFTTQSGPTQVLSITSHDLMMNDQPNAS-----MQOITTAQVS- 2712	
QY	594	GSTGAGTGNTTNTSOTVSNPTLNTYRSG-----	IDSKPTSKNKIDETNM--ADPNVTEAR 647	
Db	2713	-----GVNDLYNNRKI--	KLYEKRKSNQKSLVSNELTIQK--DQYQIYFTLPLISVANR 2763	
QY	648	IYAEYRLGID-----	NEIPIITNAGN-----FIRNTJGGVFTSGSRVYLR 689	
Db	2764	QYSKEKEIKITTSNNNPELINSNTVNSPFWASAKTOIYDNTLNATINITSATFI-----	2819	
QY	690	SYNGDQRPNTGPOFLYVGYLGYOQTKRGCTWYGTQYKLN--	NSPYDVLDSPRVGTETNQ 748	
Db	2820	-----	TYLKSDHFPYQGDIIITYLKSD 2844	
QY	749	FRRTSLTYPMWGGLPEGAR--	SFSNTPYIRAQDNPESRSIFQSGYSDNTYEY--- 802	
Db	2845	DEROEISYTTITSISSDGESEATVSNFTQDILKEAN-----	XYKLIVG 2888	
QY	803	---OSVLGFQDGIIRNNLVGYKASSFLNSRPNPNGEMIAATYLRSQIGLARTGGLPNQ	859	
Db	2889	FVRKPTLAYTINNDAN-----	NVIEDNNSNTN-----FKTILYDHKAYTVASN 2933	
QY	860	QPFCTHQVIVSIVS--PGQFSSI--	KNITLTFPGNQ-----LM-----YFLPTNEBKK 903	

Db 2934 DSINTTQTVNIDIDIGIQRTWINKKIQLVYTSSNDGEELISDQKTLIMANNHSEELSNLK 2993
OY 904 -SSVYTLRLADSSNPASSF-----SPTSLIDVNEIGVILPLIDNSF----- 945
Db 2994 HNRKKTLLKVRRIINDNKTSIIIFHLKNGIADMEYVKTKOSISISIIIEPIARAKNLOST 3053
OY 946 ---YTVNAGNVALESPSSNPSPSYAVTFNQNLS---DIAFEGSGAKYTSDFMGTIOF 999
Db 3054 QTRFELNDPDVNL---SN-----EEEAIIINGENLSVKAKKIVGS-QYLETFENNLAL 3104
OY 1000 KDEVIYIONGFTSOVARNVYTNOSFLNSLVDTTPANAGTNRKYVDPDG---NLINONLP 1056
Db 3105 MOD---TINSITFTLNK-----PIKAATN--IGINSKRIYNNFNINNP 3143
OY 1057 LKVOIQYLDGKYDAKLKNNNLVTFSPNFGALPSPVNPVPAIGSTGLIAMIILGLAIG 1116
Db 3144 LKIDNNF---SVIGPQOSDNRKRTILEYN-----KTWVKVN 3177
OY 1117 IFLRAQR-LQDKGKTFKRYDVLTPAAGSVYKKIITQTANYKK 1161
Db 3178 LELKANDNLQNLXYFIATFKDONNETV-----KSSIVSYSSIKKE 3217

RESULT 14

AB2018
hypothetical protein all1696 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2018
R:Kaneko, T.; Nakamura, Y.; Molk, C. P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MIMD:21595285; PMID:11759840
A:Accession: AB2018
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1999 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA078062.1; PID:g17135516; GSPDB:GN00179
A:Experimental source: strain PCC 7120
A:Gene: all1696

Query Match 3.5%; Score 223.5; DB 2; Length 1999;
Best local similarity 19.8%; Pred. No. 0.0013;
Matches 306; Conservative 162; Mismatches 502; Indels 577; Gaps 74;

OY 27 FCGKQSDKSNMTQ---LVNQARTLDANSVRL---AGLQNGSLFNTVLRVDN- 75
Db 278 FVFTQSGNDNTDSDADPLTGQAQIIEINPNEFNRTIDAGIYRKASLDGEVFKDANNNG 337
OY 76 -----FITANGTIKLDSTFKLYGLDSDCGGKRYKOIV-SDY----- 115
Db 338 IODAGEVGVQYVELINPTNGOVIATTT-----TNSSGYQFSGLTPGNGYQVR 387
OY 116 -----TTSNREFDOROTRA-----YVALLVNDAN-----YHLKRI----- 146
Db 388 TAPGTGIIFSTANGSNDLSDANPSTGLTQITLISGEPNGILDGVLPLASLGNFVE 447
OY 147 -NTNSRIGNRNNNSKFVIGVDNPAHYI--RTDDGKTFNFTNOGELIVDFILDAF 202
Db 448 DKNANGIQDAGETIGNATVNLDAENIATATTDNGLSFTNLRPGDYKQYFV----- 503
OY 203 ILPKDLHPDMYNIYIORKILPND-----VNTAVVWPVPGRVSGTAAODGMF----- 248
Db 504 -----QPNGFNGVSPQNGGDAIDSDGLVSDIVMLSPGE-NDTIVDAGFYTASLDGE 556
OY 249 -----DCNGGOI-----TNTDPIAOTKTTTNDQNPSTFNSGAMP----- 282
Db 557 VFEDANNNGIQDAGEVGVQYVELINPTNGDVIA---TTTNSGCGYQFSGLTPGNGYQV 613

OY 283 -----GANNRYSQNLNVKRIKITSFOLD----- 305
Db 614 RFTAPGTGIIFSTANGSNDLSDANPSTGLTQITLISGEPNGILDGVLPLASLGNFV 673
OY 306 -ERINTNSRIGNRNNNSKFVIGVDNPAHYI--RTDDGKTFNFTNOGELIVDFILDAF 362
Db 674 FEDKANGIQDAGETIGIDGATVALLDTSNGVIAATTTTDNGLSFTNLRPGDYKQYFIDP 733
OY 363 A---PILPD-----LHPDMYNI-----YIORKILP-----NDVNTAV 392
Db 734 SGFDGSPKNGVGNDAIDSDGLISDVYNLSPGENDTTVDAGFQTSASLDGPFVFNDAV--- 790
OY 393 VPMVPGRVSGTNAADGMPDCGN---GOIT-----NTDPIAOTKTTTNDQNP----- 434
Db 791 -----NNGIQDAGEIGVGVAVELINPANGVIA---TTTDANGYGFSG 834
OY 435 -----ESTFN--SGAMGANNRYSQNLNVKRIKITSFOLD----- 468
Db 835 TPEGVQKRTAPGYNFSLANGNNDALSDANVSTGITQITLISGEPNGILDGVLPL 894
OY 469 ---KFVPEMTGSEENKNTIRLATGSLPSNERHYWILDIPGPQVTLKEDSVNVSRLYL 524
Db 895 ASLGNFV-----EDKN-----ANGIQDAGEIG-----IGDQTVLLDTSNGV----- 932
OY 525 NSVNSLSFTQDSIIYIGTSELPSLWYYSF--PTRISDLTALNQVKTDIEA----- 573
Db 933 --IATTTTDGNGLYSF--TNLRPGDYKQYVQPNGFNGVSPQNGVGNDAIDSDGLISDVY 989
OY 574 -STDNGTTNG-----TTTAD-----TSSGSGTA--GT 600
Db 990 LSGENDTIVDAGFYKTASLDGEVFNVDVNNNGIQDAGEVGVASVTLTGCGADGILINGI 1049
OY 601 GNTTNSQTVSNPTLN-----TYR-----SFGIDSKRPSANKIDETWADPN-- 642
Db 1050 GDTVTYTTTNAGNNTYFAELTPGOQYQVGEVSGLPAPFOQTQANVAGNDVDS-ANPSTG 1108
OY 643 -----VIEARIYAE-----YRLGIONE--IPTN 664
Db 1109 KIQVTVLVSGENNLTLTDAGIQONAGDLSTKTDGLTNVTPGOQITTYIVARNNGLLTAVN 1168
OY 665 A--GNFIRMTIGVGSTGSGSRVYLAASVNGDQRPNGNPFLYVGYLGYYQOTRTGTW 722
Db 1169 ALVSDIIPSNLTNVTWSVASGATDNOASG---TGNINDHVTLLTGSSITTYVTGTVV 1224
OY 723 YGTYKLLNNSPYDV-LDSPRVGTE--TNOFRRTSLTPYVWGYLTFEGARSEFNT----- 774
Db 1225 SNAASGSSRRFPDONSPLDGTGNTTRFTQGTIVT-----ARAFSRVGTNG 1274
OY 775 -----PYIRAQDTPESRSIFQSGYSDNTYEVYIQVYLGFDG----- 810
Db 1275 AMSAAVLYSGYTGGLGYTDS-----EGNGGNNTHVVDNNGGRDNYVLQFSEAVVLDAK 1329
OY 811 ---IRNNLVNGVAASSF-----LNSNRPRNGLMEIATTT-----YLRSQI 848
Db 1330 LOYVLKDSIDISVIGFNPNELTNLSDSTLNSFCGEYEAANYTSSSRMADVNAAGVGNIT 1389
OY 849 GLARTSGLPNOQPFCTTHOIVSVSPGDOF-----SSIKRIRIIFPNQMLMYEL 896
Db 1390 VIALD-----TDTSPEDNFKRLRLDLVKALPABLAVMTATITTAAG--FT 1432
OY 897 FTNENKSSVYTLRLADSS-----NPDASSFSPTSLIDVNEIG----- 935
Db 1433 DTNPNNSATDWTNTIIVAPSDDLSTKTDGLTNVTPGOQITTYIVARNNGLLTAVN 1492
OY 936 VILPLDLSFTYTNAG-----NALRSSN-----PGRGSYTAVNFTNOMLSDAIE 983
Db 1493 IIPSNLTNVTWSVASGATDNOASGTGNINDHVTLLTGSSITTYVTGTVV---VSNAAST 1549
OY 984 GSGAKYTSDEWGTIOKFPDE---YLIONGFT-----SOV-ARNFVTNOSFLNS----- 1027
Db 1550 GSSTRF--DFDGSPLDGDIDGNTRFTQNGKITTAAPAFSRVDSGTNAMSAAVLYGSIYTGGL 1607
OY 1028 -LVDFTPANAGTNRVAVVDPDGNLTNQNLPKLVQIQYLDGKYDAKLK-----N 1075

Db 1608 GYTSSSENGNGNNH-VVDNNGGRDNVYLFQFSEAVVLDKAVLYLVKDSIDSWIGNEFN 1666
 QY 1076 N---NLVTFSYNNFG-----ALPSWVYPTA---IGSTLGIAI 1107
 Db 1667 NPLFVLSSTLNSFGFEYEVANTNSSSDRMADVNGVNTIVIAL 1713

RESULT 15

A64556
 toxin-like outer membrane protein HP0289 - Helicobacter pylori (strain 26695)

C.Species: Helicobacter pylori

C.Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C.Accession: A64556

R.Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKenney

son, J.D.; Kelley, J.M.; Cotton, M.D.; Meldman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser, C.

A.Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A.Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: A64556

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2893 <TOM>

A:Cross-references: GB:AE000547; GB:AE000511; NID:92313377; PIDN:AA007355.1; PID:9231338

Query Match

Best Local Similarity 3.5%; Score 223; DB 2; Length 2893;

Matches 281; Conservative 160; Mismatches 473; Indels 458; Gaps 73;

QY 14 GGLAVFG-----ALGSASFGRKOSKSDNTQVLYNOARTLDANSVRLAGQNGSL 64
 Db 287 GGTYTNEKESATNTNTAFSSGSFNFQVSSFN-GTSFNASATFQ-NQATTONSSFNQCT 344
 QY 65 F-----NVLVDNDNFITANGTIIKIDSFETKPLYGDLSDDCGKYKQIVSD- 114
 Db 345 FTTNNOTNPNNNAHQIIONSSFSFGNATLTK-----GTVNFQOAFNNS 387
 QY 115 ---YTSRNFEDQROTFRAYVALLVNDQANVHLKRTNNSRGNKNNNSKFIYGVG---- 167
 Db 388 NHQITLIONASFNNATFNNTGKITIEKDAFEN---NTFNTSVDTNNNS--VTGGVTTISG 441
 QY 168 -----DNPAHYIRFDQGTKEFNTN-----QTQGEIVNDFILDABI--LP 205
 Db 442 KNDLKNGSTIDPSSSKITLA-QCTTFNLTSISEKSVITLNSGGITTYNLNHLAINGLT 500
 QY 206 KDLHPWYLYIQKILPNDVNTAVVPMVGRVSGTNADGMEFGNGQITVTDPIAQT 265
 Db 501 SALKTN-----ESLSNPOSFAQGLMDIITYNGV-----TGQLN-ENAAISK 541
 QY 266 TTTDNDQNPSTFNSGAMPGANRRYDSOLNKHRI-KTSFOLDERINTNSRI-----GNRN 319
 Db 542 PTSSPSKSSSTNSTQV-----YQGYKIGDITTYKIQETFFSHNSIILQALBSGYT 591
 QY 320 ---NSKFEVIG-----VDNPAHVIR--FTDDGTFENFTNOGEIYND--- 358
 Db 592 PPPIYNSKIDLSASNTINADMPYDHKYYIPKSONFTESGTYLLPSVOIMGTYNSFKQ 651
 QY 359 -FIIDAPLPRDLHPM--YNLYIQKILPNDVNTAVV-----PMVGRVSGTNADGME 410
 Db 652 TFSANGSNLYIGVNTYTDHNVSSSGTVSFGDTSGSALNGHCGPWPYQCTGTT--NGTY 709
 QY 411 DC-----NGCQITN-----TDPIAQTCTTTDNDQNP-----TFNSGAMP 444
 Db 710 SAHYVITANLRSGNRIGTGAANLIFNGVDSINIANATITQHNAGIYSSMTFSQSMO 769
 QY 445 GANN---RYDSQNLVHKRIKTSQDLDEKVEYVPEWTSSEKNKTRLATGSLPSNERYI 500
 Db 770 NSQNLNGLNSNGKLSYGTFTTNEAKDKGKLFENAGQAVFENTFN----- 814
 QY 501 LDIPGTQVTLKEPSVNVFSLYLN-----SVNSLSFGDSIYIFGTSLEPLSLW 549

Db 815 ---GGSYQ--FSGDSLNSNNQFNSSGEISAKNASFNNANFNNSASFNFNSNATTSF 869
 QY 550 YSPPTRLSDLTALNQVKTDDIEASTDNGTNTGTTTADTSSGSGAGAGNT----- 603
 Db 870 VGDEFNANSNL-----QI-----AGNAVEGSTNSQNTAFNNNGVNSIGNATFDDNV 919
 QY 604 ---TNTS---QTVSNPT--LNTYRSFG-----IDSKP---TSA 630
 Db 920 ENGPTNTSVKQVYTLNNTITKLNAPLSFGDGTTFNNAHYINIAESITNGNPITLVSSS 979
 QY 631 NKIDETNMADPNVIEARIYAERYELQIONEIPITNAGNFRITRTIGVGFTSGSVLRAS 690
 Db 980 KELEYNNAKSKMLMOLINQGH--GASSBKLVSSAGNGVYDV-----YS 1022
 QY 691 YNGDPTGNQPCFL-----YFEGYLG-----YQGTTRG--TFWGYTK- 727
 Db 1023 FNNQ---TYNFOEVSQNSISIRLQVNVFEDYVMEKSDHLYQNALGFMTYMPNSYNN 1079
 QY 728 ---LLNNSPY-----DVLDSPRVGTETNQFRRTSLYFVMGQYLTREGARFSNTPYR 778
 Db 1080 NLGNANNTIYYDKSIDFYASGKTLFTKAFFSQT-----FTGQNSAIVFGAKSIWT 1130
 QY 779 AQGDTPESRSIRQSGYSNT-----YE-----YIQ----- 803
 Db 1131 SLSDAQOSTIIRFEG--DNKGAGSNDASGHCWNLOCIGFTTGHYEAOKIYITGSTEGR 1188
 QY 804 -----SVLGFDDIRN-----NLNVGVKASS--FLNSRPNPENGLEMTAATYIARS 846
 Db 1189 ISSGGGASLNFNQLGILLTNATLYNRAQTQSSSMNFYS-----NSANIOAQNSYFID 1242
 QY 847 QIGLARTSGLPNQOPGTTHQVIVSVPGQOFSSIKNIRTIIPGNQULWYLFENNNKSS- 905
 Db 1243 D--TAONGGNPN--FSEFNLNIDFNSNPFGRVGTQSVKFEKNAKNAISFTNSNLSSG 1297
 QY 906 VYTLR-----LADSNPDASSFPTSLIDVNEIGVI-----LPILDN 943
 Db 1298 LYQWQAKSYLPDMSNLVSYG---TSSIKANALINSQNASINASHSTLELOGDILVNDT 1354
 QY 944 SFY-----TVNAGNAL--FSSNPGSPGTYT---AVN-----TFENGLSDIAFEGS- 985
 Db 1355 SSLNLQSTINVSNNATINDYASLINSNGHLFENGAVNFSANITTSILNNSIYFKGAV 1414
 QY 986 --GAKY-----TSDPFGTIOFKPDEYLIONGFTSQVARNFVYNOGFLNSLVDFTPANAG 1037
 Db 1415 SLGQFNLSSNSLDPGGS-----SALTSMATFAFYDN-AFSOSPITFHQA--- 1459
 QY 1038 TNYRVVVDPCGNTLNQMLPLKVOIQYIDGKYIYDAKLNKNNLVYFSSYNNNGAL 1089
 Db 1460 LDIRAPLSLGGNLNLPN-----NSSYLD-----LKSOLV--FGDQSL 1496

Search completed: October 10, 2003, 16:36:02
 Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 11:49:11 : Search time 61 Seconds
(without alignments)
3184.940 Million cell updates/sec

Title: US-09-901-572a-4

Perfect score: 6413

Sequence: 1 MNISKKIKSYTLIGLAVFG.....APTKPTAPKPAKPTAPKE 1224

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5385	84.0	1062	AAW22721	Antigenic protein
2	507.5	7.9	1122	AAW22721	Cytadhesin protein
3	337	5.3	1144	AAW22721	Mycoplasma pirum a
4	328	5.1	1627	AAW22721	Mycoplasma pneumoniae
5	324	5.1	1627	AAW22721	M. pneumoniae pl. c
6	324	5.1	1627	AAW22721	Cytadhesin pl. c
7	252	3.9	10182	AAW22721	Staphylococcus epi
8	223	3.5	2893	AAW22721	H. pylori GHPD 148
9	223	3.5	2893	AAW22721	Helicobacter polyp

10	214	3.3	2902	22	AAW22721
11	207.5	3.2	2180	23	AAW22721
12	205.5	3.2	2180	24	AAW22721
13	205.5	3.2	2344	22	AAW22721
14	205.5	3.2	10498	22	AAW22721
15	205	3.2	2123	22	AAW22721
16	203.5	3.2	2053	22	AAW22721
17	202	3.1	2870	21	AAW22721
18	202	3.1	3178	21	AAW22721
19	199.5	3.1	1844	21	AAW22721
20	199	3.1	1612	19	AAW22721
21	198	3.1	1962	12	AAW22721
22	197.5	3.1	5795	22	AAW22721
23	197	3.1	1778	22	AAW22721
24	196.5	3.1	2150	23	AAW22721
25	196	3.1	1959	12	AAW22721
26	196	3.1	1962	12	AAW22721
27	196	3.1	1962	12	AAW22721
28	196	3.1	1962	12	AAW22721
29	196	3.1	1962	12	AAW22721
30	196	3.1	1962	12	AAW22721
31	196	3.1	1962	12	AAW22721
32	196	3.1	1974	12	AAW22721
33	193	3.0	2586	22	AAW22721
34	192.5	3.0	2261	24	AAW22721
35	192.5	3.0	2283	24	AAW22721
36	192	3.0	1536	15	AAW22721
37	192	3.0	1536	21	AAW22721
38	191	3.0	1536	18	AAW22721
39	190	3.0	2710	17	AAW22721
40	190	3.0	2710	19	AAW22721
41	189.5	3.0	1436	21	AAW22721
42	189.5	3.0	2353	17	AAW22721
43	189	2.9	1536	14	AAW22721
44	188	2.9	1072	23	AAW22721
45	188	2.9	2411	21	AAW22721

ALIGNMENTS

RESULT 1	
AAW22721	standard; Protein: 1062 AA.
ID	AAW22721
XX	
AC	AAW22721:
XX	
DT	27-FEB-1998 (first entry)
XX	
DE	Antigenic protein derived from Mycoplasma gallisepticum.
XX	
KW	Mycoplasma gallisepticum antigen; epitope; monoclonal antibody; bird;
KW	Mycoplasma Infection; Viral vaccine; avipoxvirus; herpesvirus.
XX	
OS	Mycoplasma gallisepticum.
XX	
PN	W09724370-A1.
XX	
PD	10-JUL-1997.
XX	
PF	27-DEC-1996; 96WO-JP03863.
XX	
PR	28-DEC-1995; 95JP-0352754.
XX	
PA	(JAPG) NIPPON ZEON KK.
XX	
PI	Fujisawa A, Yoshida S;
XX	
DR	WPI: 1997-363621/33.
DR	N-PSDB: AAT75087.
XX	
PT	Antigenic protein derived from Mycoplasma gallisepticum - useful in vaccines against, and for diagnosis of mycoplasma infection in birds

XX Claim 3; Pages 24-31; 45pp; Japanese.

CC This sequence represents an antigenic protein of Mycoplasma gallisepticum
 CC having an epitope specifically recognised by the monoclonal antibody
 CC (Mab) 35A6. This antigenic protein inhibits the metabolism of
 CC bird infecting mycoplasma. The antigenic protein and the DNA molecule can
 CC be used in the preparation of component and viral vaccines against
 CC mycoplasma infection in birds. Mycoplasma infection can be diagnosed by
 CC hybridising DNA in a sample with a DNA probe corresponding to 10 or more
 CC (preferably 15 to 40) bases of the antigenic protein gene.

XX
 XX Sequence 1062 AA;

Query Match 84.0%; Score 5385; DB 18; Length 1062;
 Best Local Similarity 86.0%; Pred. No. 0;
 Matches 1053; Conservative 2; Mismatches 7; Indels 162; Gaps 1;

QY 1 MNISKKLKSYTLIGLAVFAGLGSASFEGKQSDKSDNTQVNOARTLDANSYRLAGLQ 60
 Db 1 MNISKKLKSYTLIGLAVFAGLGSASFEGKQSDKSDNTQVNOARTLDANSYRLAGLQ 60
 QY 61 NGSLFNTVLDVNDNFITANGTTLIKLDSFTKFLYGLDSDCGYKVOISDYTTSTN 120
 Db 61 NGSLFNTVLDVNDNFITANGTTLIKLDSFTKFLYGLDSDCGYKVOISDYTTSTN 120
 QY 121 RFDRQRTAYYALVNDANVHLKRIINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGG 180
 Db 121 RFDRQRTAYYALVNDANVHLKRIINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGG 180
 QY 181 TKFNTQTOGEIYNDITLAPILPKDLHPDWNTLYIQRKILPNDVNTAVVPWVGVSQ 240
 Db 181 TKFNTQTOGEIYNDITLAPILPKDLHPDWNTLYIQRKILPNDVNTAVVPWVGVSQ 240
 QY 145 ----- 144
 Db 145 ----- 144
 QY 241 TNADGMEFCGQITNTDPIAQTITTNDQNPSTFNSGAMPGANNRYSQLVKHKRIKT 300
 Db 241 TNADGMEFCGQITNTDPIAQTITTNDQNPSTFNSGAMPGANNRYSQLVKHKRIKT 300
 QY 145 ----- 144
 Db 145 ----- 144
 QY 301 SFQDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGGTKFNTQTOGEIYNDFI 360
 Db 145 -----RINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGGTKFNTQTOGEIYNDFI 198
 QY 361 LDAPILPKDLHPDWNTLYIQRKILPNDVNTAVVPWVGVSQTNADGMEFCGQITNT 420
 Db 199 LDAPILPKDLHPDWNTLYIQRKILPNDVNTAVVPWVGVSQTNADGMEFCGQITNT 258
 QY 421 DPIAQTITTNDQNPSTFNSGAMPGANNRYSQLVKHKRIKTQFQDEKRYVPEWTSSE 480
 Db 259 DPIAQTITTNDQNPSTFNSGAMPGANNRYSQLVKHKRIKTQFQDEKRYVPEWTSSE 318
 QY 481 NKNTIRLATGSLPSEREWIILDIPGTPVTLKEDSVNVFSRLYLNVSLSFIQDSIYIF 540
 Db 319 NKNTIRLATGSLPSEREWIILDIPGTPVTLKEDSVNVFSRLYLNVSLSFIQDSIYIF 378
 QY 541 GTSLEPLSMYSEFPRLSLDTALNQVTKTDIEASTDNGTTTNGTTTADTSSSTGAGT 600
 Db 379 GTSLEPLSMYSEFPRLSLDTALNQVTKTDIEASTDNGTTTNGTTTADTSSSTGAGT 438
 QY 601 GNTTTSQTVSNPTLNTYRSRGIDSKPTSAKIDETNADPNVIEARIYAEYRLGIONEI 660
 Db 439 GNTTTSQTVSNPTLNTYRSRGIDSKPTSAKIDETNADPNVIEARIYAEYRLGIONEI 498
 QY 661 PITNAGNFIRNTIGVGFTSTGSRVLRASVNGDGPPTGNFOPELYVFGYGYOQTRTGT 720
 Db 499 PITNAGNFIRNTIGVGFTSTGSRVLRASVNGDGPPTGNFOPELYVFGYGYOQTRTGT 558
 QY 721 FWGTYKLLNSPYDLSRPVGTETNOFRRTSLYPVAGVLEEGARSPNPFYIAQ 780
 Db 559 FWGTYKLLNSPYDLSRPVGTETNOFRRTSLYPVAGVLEEGARSPNPFYIAQ 618
 QY 781 GDPEPSRSIFQSGYSNTEYEQSVLGEFDGIRNNINLVGKASSFLNSNRPNGLEMTAA 840
 Db 619 GDPEPSRSIFQSGYSNTEYEQSVLGEFDGIRNNINLVGKASSFLNSNRPNGLEMTAA 678

QY 841 TTYLRSGIGARTSGLPNOQPFQGTTHQVTSVSPGDQFSSIKNIRITIFPGNOLMYEFLTNE 900
 Db 679 TTYLRSGIGARTSGLPNOQPFQGTTHQVTSVSPGDQFSSIKNIRITIFPGNOLMYEFLTNE 738
 QY 901 NKSQSVYTLRLADSSNDPDASSSPSLIDVNEIGVILPLDNSFYTVNAGNVALFSSN 960
 Db 739 NKSQSVYTLRLADSSNDPDASSSPSLIDVNEIGVILPLDNSFYTVNAGNVALFSSN 798
 QY 961 PGSPGSYTAVNTFNQNLSDIAFEGSGAKYTSDDWGTQFQKPDXYLIONGFTSOVARNFT 1020
 Db 799 PGSPGSYTAVNTFNQNLSDIAFEGSGAKYTSDDWGTQFQKPDXYLIONGFTSOVARNFT 858
 QY 1021 NOSFLNSLVDFTPANAGTNRVYVDPDGNLTNQLPLKVOIOYLDGKRYDAKLNNLVY 1080
 Db 859 NOSFLNSLVDFTPANAGTNRVYVDPDGNLTNQLPLKVOIOYLDGKRYDAKLNNLVY 918
 QY 1081 FSYNNFGLPSPVVPYTAIGSTGLILAIMIILGLAIGIPLRAQRKLDGKRTFFKRYDVL 1140
 Db 919 FSYNNFGLPSPVVPYTAIGSTGLILAIMIILGLAIGIPLRAQRKLDGKRTFFKRYDVL 978
 QY 1141 TAAVGSYVKRTITOTANVKKRPAALGAGSGDKKPAAPAKPSAPKASSPAKPTG 1200
 Db 979 TAAVGSYVKRTITOTANVKKRPAALGAGSGDKKPAAPAKPSAPKASSPAKPTG 1038
 QY 1201 PKSGAPRTAPKPAAPKPTAPKE 1224
 Db 1039 PKSGAPRTAPKPAAPKPTAPKQ 1062

RESULT 2
 AAR64927
 ID AAR64927 standard; Protein; 1122 AA.
 XX
 AC AAR64927;
 DT 10-AUG-1995 (first entry)
 DE Cytadhesin protein.
 XX
 DE Cytadhesin protein.
 XX
 KW Cytadhesin protein; Immunoassay; antigen.
 XX
 OS Mycoplasma gallisepticum.
 XX
 PN US378820-A.
 XX
 PD 03-JAN-1995.
 XX
 PE 09-NOV-1992; 92US-0973257.
 XX
 PR 09-NOV-1992; 92US-0973257.
 XX
 PA (DOHM/) DOHMS J E.
 XX
 PA (KEEL/) KEELER C L.
 XX
 PI Dohms JE, Keeler CL;
 XX
 DR WPI; 1995-051314/07.
 DR N-PSDB: AA081778.
 XX
 PT Nucleic acid encoding cytodhesin protein - used as a probe to
 PT diagnose Mycoplasma gallisepticum infection in poultry
 XX
 PS Claim 1; Column 7-16; 13pp; English.
 XX
 CC The protein is a cytodhesin protein from Mycoplasma gallisepticum,
 CC and may be used as a diagnostic antigen (optionally in recombinant
 CC form) in immunoassay formats for diagnosis of M. gallisepticum
 CC infection in poultry, e.g. turkey or fowl.
 XX
 SQ Sequence 1122 AA;

Query Match 7.9%; Score 507.5; DB 16; Length 1122;

indicated in the Features Table are claimed; they inhibit infection of human cells by M. plium and infection of human T lymphocytes by CC HIV.
(updated on 25-MAR-2003 to correct PN field.)

Sequence 1144 AA;
Query Match 5.3%; Score 337; DB 16; Length 1144;
Best Local Similarity 21.5%; Pred No. 2,2e-12;
Matches 299; Conservative 168; Mismatches 491; Indels 434; Gaps 74;

```

OY 5 KRLK---SYTLIGLAVFGLGASGFGKQSDKSDNTQOLVNO-----ATLTANSLRL 55
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 2 KRLKREYKLLLSLSTIVSAAISLSTFNKDOIINQNVKSFNSPNSIVNKKYK 61
OY 56 AGIGONGSLFNLRYVDNFTTANGTILKIDSPKPLYGLD-----LSDD-----CGGY 106
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 62 IRHMOMNN--ENGVEIKNGSFVVLTSQSATRIDAGNIIEMFDPKIASSESOYANLAG 120
OY 107 KVKQIVSDYTTSHRFDOQRATYALLV-----NDEANVHLKRIINTNSNRIGNNNNSK 161
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 121 KVEITIQDEBDNEN-----ILYLLIPKNTPPKQASIDPKDLAYAVNELGSSKSKQ 172
OY 162 FVT-----GGVDNPAHYIR--FTDDGTFK--NFTNQTGEIYVDFILDAPILPK 206
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 173 ATVVOIIEENVLYQSGSTWPSFTIKGLMHIDPKKMDYIPNOMKSSQS-----STFFTK 227
OY 207 DLHPDWNYLYIQKRIIPN--DVNT-----AVVPW-----VGRVSGTNAADGMD 250
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 228 EDHSWYVANNSEKVRHADQNTQNGKLKSNAMYLPWKQYITNIGNMFAKGIYILFQ-- 286
OY 251 NGQITNTDPIAQT-----KTTNDONPSTFNSGAMPGA-----NNRYDSOLNV 294
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 287 NGGSIYN--DEPAISIGMKLDFLKPYSGNIDNO-----NYGGIPYLYLRLRYDPS-- 338
OY 295 KHRKITSFOLDERINTNSNRIGNNNNSKFEIGVNDPAHYIRFTDGTKEFNTQOGE 354
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 339 KPLIGTS-----APNR--RMNQSYAPIGQTNFTYVPLAVGQVQAINASD-- 382
OY 355 IVNDFILDAPILPKDLHPDWNYLY-----ORKIIPNDVNTAVVPWPVGRYS 401
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 383 -----EATYLLAAGITVGAKESQAREVISNS--NT-----S 413
OY 402 GTNADGMEFDCNGQOITNTDPIAQTKTITTDNQNSTFNSGAMPANNRDSDQLNVKRRK 461
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 414 TNKVYTKIQDKRSIQIGLANTITNTKDT-----AANSIDPAL-----LFG 453
OY 462 TSFQIDKEFYVPEWTGSEKNTIRLATGSLPSNERWIIDPQTQVTLKEDSVNVSFR 521
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 454 TAFNIDSLINLPT--TKLINENLTFQ--NVFOYESIF--DVGATMVSASAVGTYYYFDK 506
OY 522 -----LYLNSVSLFIDSIYIFGTSLEPLWYSEPTRLSDITLALNOYKTDIEASST 576
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 507 KHHASSSTTDINTYTTASNWNNGNTRAF--MSYK--PN--NDIGSIFQKIND----- 555
OY 577 DNGTTNGTTTADTSSGAGNTNTNTSQYVSNPLNMY--RSEFGIDSKPTSA--NK 632
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 556 NNNAAT-----YSTNLLELLEN--AINYYTSLSPGISLKLGLGLTK 594
OY 633 IDETWADPNVLEARIYAERLGIONEIPTNAGNFINTIGVGFSTGSRVYLARASYN 692
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 595 IEMPSKEN-----ENTLYGYAM--QVGSKIYVYLNPK 625
OY 693 GDQPTGNFQPLLYFGYLGIOQTGTGFWG-----TYKLANSYVDLDSPRGTENQ 748
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 626 SDLSIAVHGSPSISIG--ESNLVGSAGKYGDMYPVYKINNSIGIVPSD--YSNTNN 680
OY 749 FRTSLTYVMGGLYLEEGARSESN--PIRAQCDTPESRSIFOSGYSDN-----TYEYI 802
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 681 IINTGVAL-----YVF--GIKDPNDIPIIASQFE-----IGNSPYEDNSSTKTNGTL 727
OY 803 QSVL--GFGDGIKNNINLVGVKASS--FLNSNRPNGLEMAIATYTLRSQIGLARTSGLPN 858
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|

```

```

Db 728 OPTIPNDVEYGL--NSTNPNSEISSLMLNNQTKTNNE----- 764
OY 859 QQPFETHQVIVSGDDPSSKTKRTITFPNQLM-----YFLTENNEN-----KSSVYT 908
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 765 -----HEIVTKSP--EISE-----YVGAIAWTEREFYNYSSNNADKSKRAMFE 808
OY 909 IRLADSSNDPDASSFPSFLIDVNEIGVILPLDINSFYTV-----N 949
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 809 VK-----DSENLSNSETTYGMQVGLDNLTAQSYVQKNNRQPGDFVLLRTFDDT 860
OY 950 AAGNVALFS--SNPQSPG--SYTAVNTFNQNLSDIFESGAKYISDFEGCTIQFPRDEYLI 1006
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 861 OKNNDIFFEQIINTREPGISYCKL--KQN-----YGSFYETISEIDRLSL 905
OY 1007 QNG--FTSQVARNVYTN-----QSFNLNVDFT--PA 1034
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 906 GNGQFVNNLSNQIITNLNLANLVQVNLSTVTGNPRLDSKSTIRIVKNOFLNEVQVTKNPI 965
OY 1035 NAGTNVR--VVVDPDGNLTQNLPLKVOIYLDGKIYDAKLNN-----NLVTFES--YNN 1085
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 966 IEGSTPYGVVILVASNVDFVSQATFTAYSNNSLTKNDVMPRTNSSKINVINNSIRAG 1025
OY 1086 FGALPFWVPYFALGSTGLIATIMTILGAIPIRLAQRKLDQGFRTFKVYDLTLAAG 1145
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 1026 FSAMADWILPVYLAIPVYLAIVLITGLGCSIGIPMAKHKAKIIVGFELQDKVGLTSAVG 1085
OY 1146 SVYKKII--TQTNANVKKRPAALGAGSGDKRPAALAKP--AAPAKGAPAKASPAKPTGP 1202
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 1086 GVFKIITDNTSNNSVSKPQML-----KAAAKRPVTPPARSQTLTNDVSRTPPS 1136
OY 1203 SGAPTKPTAPK 1214
    ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:| ||:|
Db 1137 S-----APRP 1141

```

RESULT 4
AAR47911
ID AAR47911 standard; Protein: 1627 AA.

AAR47911;
25-MAR-2003 (updated)
28-JUL-1994 (first entry)

Mycoplasma pneumoniae cytoadhesin P1 protein.
Mycoplasma pneumoniae; P1 cytoadhesin; antigen; immunogen; epitope;
anti-mycoplasma vaccine; pathogen.
Mycoplasma pneumoniae.

Key Location/Qualifiers
Peptide /label=1383..1395
/note="Claimed fragment of P1 protein, involved in cytoadherence"

Region /label=1356..1395
/note="Claimed fragment of P1 protein, involved in cytoadherence"

Region /label=1383..1518
/label=epitope
/note="Claimed fragment of P1 protein, involved in cytoadherence"

Region /label=1401..1518
/label=epitope
/note="Claimed fragment of P1 protein, involved in cytoadherence"

US5281694-A.
25-JAN-1994.

FT /label= pref.-fragment
 XX US502636-A.
 XX 25-JUN-1991.
 XX 10-NOV-1987; 87US-0118967.
 XX 10-NOV-1987; 87US-0118967.
 XX (TEXA) UNIV OF TEXAS BOARD.
 XX Baseman JB, Su CJ, Dallo SF;
 XX WPI: 1991-207465/28.
 XX N-PSDB: AAQ12518-20.
 XX Mycoplasma pneumoniae P1 protein and gene - used for studying
 XX mycoplasma infection and for producing diagnostic reagents and
 XX vaccines
 XX Disclosure: Fig 6(A-N); 39pp; English.
 XX Amino acid rpp is encoded by either TGG or TGA.
 XX The peptides can be used for producing antibodies for e.g.
 XX diagnosis. The MP P1 protein has a mol. wt. of 160-175 kD.
 XX and has a cytohesin epitope.
 XX See also AAQ12518-20.
 XX SQ Sequence 1627 AA;
 Query Match 5.1%; Score 324; DB 12; Length 1627;
 Best Local Similarity 21.0%; Pred. No. 2,4e-11;
 Matches 267; Conservative 146; Mismatches 398; Indels 458; Gaps 68;
 251 GNGQTT---NDPI-----AQTCTTDNONPSTFNSG 279
 506 GNGHYTKSAHTAPLSIGFVRVYNATGSAVTGMPYALLFSAVNKOTGLKDLDPFN-- 563
 280 AAMPANRRYSQNLVNRKIKTSFOLDERINTNSNRGNNSKFFVIGC---VDNPAHVI 336
 564 -----NRRV-----FEYVPMVAAGAKFYGR---ELVLAGTTTMDGATVP 601
 337 RETDD--GTRKFTNOTGELYNDEIL-----DAPI-----LPKDLA- 371
 602 RLLYDELESNLVAGGGLKREDQLTPPYGMANRPDLPGAMSSSSSSSHNAPYYFHN 661
 372 -PDVNYLYIQRKILPNDVNTAVVPPVGVSGTANAD-----GMPCCGGOITN- 419
 662 NPDMDRPIQ-----NVYDAFIKPEW---DKNGKDDAKYIYPRYSGMW---AMOVYINW 709
 420 TDDIACOTKTTD--NONPSTFNS-----GAMP-----GANN-----RYDSQ 453
 710 SNKLTDOPLSADVENAVQPNLSFALLINPELLALAPDKVYKGENEFANNEYERENOK 769
 454 LNVK-----HRIKTSFO-----LDEKFFVPEWGTGSEENKNIPLATGLSPS 494
 770 LTVAPPGCTMWSHSPFLSRFTSGFNLVGSVLDQVLDVYPMWIGNGIRG-----N 819
 495 NERWILIDIGTPQVYLKEDSVAVFSRLYNSVNSLSFI-----GD 535
 820 NHR-----GVDDITAPQTSAGSSSGISTINTSGSFLPTSPNIGVGLKANVQATLIGS 872
 536 SIYFGTSELPLSLVYFPTRLSDTALANQVKT-----DDIEASSTDNCTTNGTTTAD 590
 873 QTMITGGS-----PRRTLDQANL-QLMTGAGWRNDKASSGSGSDENHTRFISATGM 921
 591 TSSGSTGAGTNTNTSQ-----TVSNPLNTYRSRIGDSKPTSAKIKDETINWAD--PNVIE 645
 922 DOOGSGTSGAGNPPLSKODNISKSGSLTTQDGNALDOQ-----EATNYTLPPLNLP 974
 646 ARIVAYELGIGNEIPIITNAGN-----FIRNIGGVGFTSTGSRVYLRSYIGDORPTG 699

DB 975 T---ADW-----PNAISFTKNNNAQRAQLFLRGLLGSI-----PVLVNRSGSDSNKRFQA 1020
 QY 700 NFQPELYVGYLYGQOFTGTGFWYGYK-LLN-----NSPFDVDSPRVG- 743
 DB 1021 TDQKWSYF--DLHSDQTKLNLPAVGEVNGLLPALVETVFGNTRAGSGSNTTSSPGICF 1078
 QY 744 --TETNORRRISLYTP-----VMGYYLTEEGARSFNTPYIRAO 780
 DB 1079 KIPQONDSKATLITPGIAMTPDOVGNLYVSGTIVSPQOLGWLVL-----FTFDVFKPR 1131
 QY 781 GD-----TPESRSIFQSGSDNTYTXIOSVGLPDI-----RNNL 815
 DB 1132 AGYLGLQLGLDASDATORALIMAPRPAWAFRGSNV-NRIGRVSDVLDKGVNADQASD 1190
 QY 816 NVGVKASSFLNSNRPNPNTL-----EMIAATYTL-RSQIGLARTSGL----- 836
 DB 1191 SOGSTTAVRNALPEHPNALAFQVSVLASAKKPTSSGQGTQRTNSSPYLLVLPKRYQ 1250
 QY 857 -----PNO-----OPFGTHQVIVSPGDQFSSIKNIRTIFFGNOLWYELF 897
 DB 1251 SDKLDDDLKNLDPNQRTRKLQSFQTDH---STQPOFO--SLKTTTFVFG----- 1296
 QY 898 TRENKSSVYTLRLADSSNPDASSFPSLIDVNEI-----GVLLPLDINSFYTVNAAGN 953
 DB 1297 TSSGNLSVLS-----GGAGGSSSGSGGVDLSPVEKSGMLVGQLPST-----SDGN 1346
 QY 954 VALFSSNPGSPGTYAVNTFNO-----NLSDAFESSGAKYSDPMGTIOFKPDEYLION 1008
 DB 1347 TS--STNNLAPNT---MTGNDVYVGVRUS---ESNAKMDVDGIVRPLAELLDGE 1396
 QY 1009 GFTSQVARNFVT-----NOSFLNSLVD--FTPAN-----AGNRYV- 1042
 DB 1397 GQTDADTGPOSVKFKSPQIDFNRLFTHPYTDLPDPTMLVYDQYIPLFDIPASVNPKNV 1456
 QY 1043 ---VVDPDGNLTNOMLPLKAVQIOTLDGKYTDALKNN-----NLVYFS----- 1082
 DB 1457 RLKVLSPDTN--EQSLGRLFEFFKPD---ODTPNNVVOVNPNGPFLPLTASSGQPT 1511
 QY 1083 -YNNFGALPSWVVPYTAIGSTGLIAMIIGLAIGIPLRAORKLQDKGRTTKKVDILT 1141
 DB 1512 LESPENQMPDYVLPPLAITVPIVYLVSTLGLAIGIMHRKQALAGFALSNOKVAVLT 1571
 QY 1142 AAVGSVYKRIITQGANVKKRPALGAGKSGDK-KPAAAPAPAPKPSAPKASSPAKPTG 1200
 DB 1572 KAVGSVFKFEIINRT-----GISQAAKRLKQTSAAKPGAPRPVPPKAPKRPVQ 1621
 QY 1201 PKSGAPTKP 1209
 DB 1622 P---PKRP 1626

RESULT 6
 AAR67538
 ID AAR67538 standard; Protein: 1627 AA.
 AC AAR67538;
 XX 25-MAR-2003 (updated)
 XX 07-AUG-1995 (first entry)
 DE Cytohesin P1.
 XX P1 protein; cytohesin; mycoplasma; hybridization; probe.
 XX Mycobacterium pneumoniae strain M129.
 XX Key Location/Qualifiers
 XX Peptide 1..59
 XX /label= sig-peptide
 XX US5369005-A.
 XX 29-NOV-1994.

XX (GENO-) GENOME THERAPEUTICS CORP.
 PA Doucette-Stamm LA, Bush D;
 PI WPI: 2002-381255/41.
 DR N-PSDB: ABN90859.
 XX Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 PS disclosure; SEQ ID 3159; 267pp; English.
 XX
 XX ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly *S. epidermidis* infections. The sequences can be used to
 CC screen for compounds able to interfere with the *S. epidermidis* life
 CC cycle or inhibit *S. epidermidis* infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

Sequence 10182 AA;

3.9%; Score 252; DB 23; Length 10182;

Query Match Best Local Similarity 20.44; Pred. No. 1.1e-05;
 Matches 284; Conservative 148; Mismatches 407; Indels 556; Gaps 73;

31 QSDKSNNTQLVNQAFLDANSVRLAGLQNGSLFTVLDVNDFTTANGTITKIDSF 90
 1398 QATKANSRLR-----SNINSNEKQLAERGSGSKSLTRDDGKSLTNSN----- 1443
 91 TKPLXGLDSDCGGYKVKQIVSDYTSRRRFQROTFRAYALLVDEANVHLKRIITNS 150
 1444 --PIQVLDLVEPDNGYGRQY-----SHS 1465
 151 NRIGNRNNNSKFEVIGVDNPAHVIRFT-DGSTRKNTNQ--TQGEIVNDFILDAPIPKDL 208
 1466 NVIYNEKNS--IVNGQVPEANGASAFNIDKYKMANANGIMGYITAAQIXLAPSPKG- 1523
 209 HEDWNLVYOR-----KILPND-VNTAVVPPVGR-----VSG-----T 241
 1524 -----YIEKLGQMLSTNNVINYFVPSDKVPSIT--VGYNDHHFYVSGTEKNTI 1573
 242 NADGWFDCNGQIINTDPIAQTCTTDN-----QNPSTFNS-----GAMPGANRRYD 289
 1574 NVND--NYGLNTVASTDSATITMTNNNELVGAPNVTNSINKIVKKATDKGNSIV 1630
 290 S-----QANVHKRIKTSFOLDERINTNSNRIGNRNNNSKFEVIGVDNPAHVIRFTDD-G 342
 1631 SFTVNIRKPLNEKRITITSS-----NQPVRISNIGNNANSTIEDQNRVKSLSMTKILG 1685
 343 TKFNTNQTQGEIVNDFILDAPIPKDLHPDWNLYIQKILPNVNTAVVPPVGRYSG 402
 1686 TR-NYVNESN-----NDVRSQVVS-KVNR-SG 1709
 403 TNADGWFDCNGQIINTDPIAQTCTTDN-----QNPSTFNS-----GAMPGANRRYD 452
 1710 NNAIVNTTTFSSGTNTTIVPYKHAVLLEVFPTRITTVAGQ--FPGKGTSPPDIFS- 1765
 453 QANVHKRIKTSFOLDERIVPEWTGSEEN-----KNITRLA 488
 1766 -----LRTGGPDAIVWVNNQPDINSNOIGRDLTLHAIFEDGETPIRKDTYKL 1818
 489 TGSLLPSN-----ERYW--TLDI-----PSTPO----- 508
 1819 SOSIPIKQIYETTINGREFNSGDVAPGNFVOAVQIYPEHMDPFMAGSGSTPSRRNAGSFT 1878
 509 --VTL-----KEDSVNFESRLYLSN--VNSLSFT-----GDSITYFGTSELPSLWYYSF 553

1879 KTVTVYQNGCTENVANVLEKVPKPKVYIDNSVYSKQNLNQQOILVRNVPQNAQVTLXQ- 1937
 554 PTRLSDDLALNQVKTDDIEASSTDNGTGTGTTADTSSGSN-----GAGTGNNTNSQ 608
 1938 -----SNGTVIPNTNTIDSNGLAIATVTTQGLTPTGNTITAKTS 1974
 609 TVSNPTLNTYRSFGIDSKPT-----SANKIDET-----KNADPN 642
 1975 MTNNVTYTKQNSSGIASNTTDDISVSENSQVNVTAGMAKNDGIKIKGTNYNPNDRN 2034
 643 VLEARIVAEYRLGIONE-----IPIT-----NAGNF 668
 2035 SFTSNIPASHSLTWNEEBSNWKNNIGTTKYTVTLPLNHOGTFRVDPILTIYPTVAKNP 2094
 669 IR-----NTIGVGFTSTGSRVYLRSYNGDQRTGNF---QP 703
 2095 VHDQGRNLNTQTDVYNIIFENNRLAGT-----ASWKNNRQPDKNAGVON 2142
 704 FLYVEGYLG-----YQGTTRGT-----WGYTKILNNS---PY 734
 2143 LIALVYPIGSIPLLEVYKVVVNFDPQPIYKIQVGDTPFKGTWAGYKHLNGBGLPI 2202
 735 D-----VLDSPRVGTETNQ-----FRRTSL---TYPVN-----GYLTGEGAR-----SF 771
 2203 DGMKFTYNNQOSTGTTSDQMSLAYRTTPEVKGTQDVVNPNSMGWQTSQAKFIVTNAK 2262
 772 SNTPIYIRAO--GD--TPES-RSIFQSGYSDMTVEYIOSVGFQDGRNNLWGVKASFL 825
 2263 PNPPTTOSKGTQGVTVTPGAVNRILISG-----TNDIQA--SADKIIVLNK--GNKLTTFV 2315
 826 NSN-----RPNNGLEMIATYILRSQIGLARTSGLPNQOPFGTHOIVSYSPQDQ 876
 2316 KNDGKRTVETGSPDINGI-----GPTNNGTALSLSR-----LAVRPDQ- 2354
 877 FSSIKNIRITFPGNQLMYFLFTENKSSVYTLRLADSSN-----PDAS-S 921
 2355 --STEATATGSGSETI-----STASSETIYKAPQPOQVATHHYDNGTPIILDNSRN 2406
 922 SEFPTSLIDVNEIGVILPLDLSFYYVNAAGNVA--LFSNPGSPGSIYAVN--TFN 974
 2407 SLNPTREYEN-----YTEKLNGNEFTOKSFYITKNNKMKWINKPNVYERN 2453
 975 QNLSDIAEFGSAGARTSDFMGTIOFKRDEYLLNGFTSQVARNFVGNOSFLNSLVDFPA 1034
 2454 QDNKGVVSANTIKPNSOI--TTPKAGQGTEN-----INPIYQA-----PA 2495
 1035 NAGTYRVYVDPDG--NLFTNQLPLKVOIQYLDGKYDALKNNNLVFTSTNNFGALPSWV 1093
 2496 QHTLTINELVKEQOGVYINDINNAYQV-----PNKRVAIKQGN--ALPT-- 2539
 1094 VPTAIGSTLGLALM 1108
 2540 -NLAGGSTSHLPVVI 2553

RESULT 8
 AAM98828
 ID AAM98828 standard; Protein: 2893 AA.
 XX AAM98828;
 XX 31-MAR-1999 (first entry)
 DE H. pylori GHPO 1484 protein.
 XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 XX Helicobacter pylori.
 OS WO9843478-A1.
 XX 08-OCT-1998.

XX 01-APR-1998: 98WO-US06371.
PF
XX 29-JUL-1997: 97US-0902615.
PR
XX 01-APR-1997: 97US-083457.
PR
XX 24-JUN-1997: 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI
XX A1-Garawi A, Kleanthous H, Miller C, Coenen RP, Tomb J;
XX WPI: 1998-542293/46.
DR N-PSDB; AAX14547.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
XX Claim 8: Page 1827-1840; 2054pp; English.
XX
XX This sequence represents a Helicobacter pylori GHPD protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
XX Sequence 2893 AA:
SO
Query Match 3.5%; Score 223; DB 19; Length 2893;
Best Local Similarity 20.5%; Pred. No. 0.00013;
Matches 281; Conservative 160; Mismatches 473; Indels 458; Gaps 73;
QY 14 GGLAVEG-----ALGSAPFGKQSDKSDNTOLVNOARTLDANSYRLAGLGGSL 64
DB 287 GGTYFNKEPSATNNTAFSGSEFNKGVSSFN-GTFSNMSYTFD-NQATFQMSFNGGT 344
QY 65 F-----NTYLKRVDDNFITRANGTIKLDSTFKPLKGLDLDSCGGYKQVSD- 114
DB 345 FTFNNQTNPTNNAHQDIONSFSFGNATTLK-----GFVNFOQAFNNS 387
QY 115 ---YTSRNRFDQRFRAYVALVNDENAVHLKRITNSNRIGNRNNSFEVIGV- 167
DB 388 NHQLTIQNASFNATNNTNNGKITIEKDAFN---NTFNTSVDTNMS--VTGGVTLSC 441
QY 168 -----DNPAHVIREFDGTGKFNFTN-----QTGGELVDFIIDAPT--LP 205
DB 442 KNDLKGSTLDFGSSKITLA-QGTYFNLSLSEKSVTLNLSGGITYSNLHNLINGLT 500
QY 206 KDLHPWYNLYIQRKILPNDVNTAVPWPVGRVSGTNADDMGDCNGQITNTDPIAQR 265
DB 501 SAKLTN-----ESLSNPQSFAGLMDITLYNGV-----TGGQLN-ENATSK 541
QY 266 TTTDNQNPSTFNSGAPGANNRYSOLNKHRI-KTSFQDERINTNSNRI-----GNRN 319
DB 542 PTDSPPSKSSTNTQV-----YQCYKRGDIIYKIQEIFFSHNSIITQALSGSTY 591
QY 320 -----NNSKFIYIG-----VDNPAHVIR--FTDDGTYKFNFTNQTGEIVND-- 358
DB 592 PPPVINGSKFDLSASVYNADMPWYDHKYYIPKQNFTESGTYLPSVQIMSGSYNSFKQ 651
QY 359 -FLIDAPILPKDLHPW--YNLYIQRKILPNDVNTAV-----PWPVGRVSGTNADDMG 410
DB 652 TTSANGSNLYIGYNTWTDHNVSSGTVSFGDTSGSALNCHGCPWPVYQCTGT--NGTY 709
QY 411 DC-----GNQOITN-----TDPIAQRKTPTDNQNPST-----TFNSGAMP 444
DB 710 SAVHYITANLBSGNIGTGGANLIFNGVDSTINANATITQONAGIYSSSMFSTQSM 769
QY 445 GANN-----RYDSOLNKHRIKTSFQDERKFEVPEWGTGSEBNKNIITRLANGSLPSNERYWI 500

DB 770 NSQNLNLSNCKLSVYGTTFTEAKDKPIFNAGQAVFENTNEN----- 814
QY 501 LDIPGTPOYTLAKDSVNVFSRLYN-----SVNSLSTIGSIIYFGSELSPLM 549
DB 815 ---GGSYQ--FSGDSDINFSNNQFNNGSGFEISAKNMSFNANFNNSASFNFNNSNATTSF 869
QY 550 YSFEPTRLSDTLALNOKVDIIEASSTDNCTTNGTTTADTSSTGAGCTGT----- 603
DB 870 VGDFTNANSNL-----OI-----AGNAVFNSTNGSONTANFNNTGVSINISGATPDNVY 919
QY 604 ---TNTS-----QTVSNPT--LNTYRSEF-----IDSKP-----TSA 630
DB 920 FNGPTNTSVKGOVTLNNTITLKNLNAFLSEDCGTFITNAHSVINAESITNGNITLVSSS 979
QY 631 NKIDETNMADPNVIEARITAEYRLGIONEIPITNAGNFIRNTIGVGETSGRYYLRAS 690
DB 980 KEIEYNNARSKNLMOLINOGH--GASSEKLVSAGNGYDVY-----YS 1022
QY 691 YNGDQRPRTGNFOFL-----YVEGYLG-----YQQRRTG--TFWYGYK- 727
DB 1023 FNNQ---TYNFOVEFSQNSISIRLGVNMYFDYVDEKSDHLIYONALGPMITMPSYN 1079
QY 728 ---LLNNSPY-----DYLDSPRVGTETNOFRRTSLTPVWGYLTEEGARSPSNPYIR 778
DB 1080 NLGNANNTIYYOKSIDEVASGKTLPFKAEFSQT-----FTGONSAIVGAKSIWT 1130
QY 779 AGQDTPESKSIPOSGISDNT-----YE---YIQ----- 803
DB 1131 SLSDAPQSNITIFFG--DNKAGSNDASGCHWNLOIGFTTGHYEAKIYITGISIEGMR 1188
QY 804 -----SVLGFDCIRN-----NLNVGKASS--FLNSNRPNPNCLEMIATYLRAS 846
DB 1189 ISSGGASLNFENLQGLILTLNATLYNRACTOSSANFTS-----NSANIQONSTFYD 1242
QY 847 QIGLARTSGLPNOQPGTTHQVIVSPGQFSSIKINIRLTPGNQMLYFLFTENKKS- 905
DB 1243 D--TAQNGNPN--FSPNALNDPNSSPRGVYGTQSVFKFNKNAIAISFTNSTNLSC 1297
QY 906 VYTLR-----LADSSNDASSFSPTSLIDVNEIGV- -LPLDN 943
DB 1298 LYOMQAKSVLEDNSNLSVSG--TSSIKANALNLSOMASINASNSTLELQDLDLVNDR 1354
QY 944 SFY-----TYNAGANVAL--FSSNPSPGSGSYT--AVN-----TFNQLSDIAFEGS- 985
DB 1355 SSALNOSTITVNSNNTIINDYASLINSNGHLNFCNAVFNNSANITTSLNSSIVYRGAV 1414
QY 986 -GAKY-----TSDFWGTIQKRPDEXYLLQNGFTSOVARNFTNOSFLNSLVDFTPANAG 1037
DB 1415 SLGQFNLSNNSLDFQGS-----SAITSMYAEFYPDN-AFSQSPITEFOA-- 1459
QY 1038 TNYRVVVDPPGNLTNQNLPKVOLOYLDGKYVDAKLNNMLVFFSYNNFGAL 1089
DB 1460 LDIKAPLUSLGGNLLPNV-----NSSVLD-----LKSOLV--FEDQDSL 1496
RESULT 9
AAW71556
ID AAW71556 standard; Protein: 2893 AA.
XX AAW71556;
XX
XX 09-NOV-1998 (first entry)
XX
XX Helicobacter polypeptide GHPD 1484.
XX GHPD 1484; infection; therapy; diagnosis; vaccine; gastritis;
XX ulcer.
XX Helicobacter pylori.
XX W09821225-A1.
XX
XX 22-MAY-1998.

XX 14-NOV-1997: 97WO-US21353.
XX 28-JUL-1997: 97US-0902615.
PR 14-NOV-1996: 96US-0749051.
PR 01-APR-1997: 97US-0831309.
PR 01-APR-1997: 97US-0833457.
PR 01-APR-1997: 97US-0834705.
PR 24-JUN-1997: 97US-0881227.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (PLAC-) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR-) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Haas R, Kleanthous H, Meyer T, Miller C;
PI Odenbreit S, Tomb J;
XX MPI: 1998-297855/26.
DR N-PSDB; AAV52091.
XX Helicobacter pylori polypeptide and polypeptide sequences - useful to
PT treat or prevent gastrointestinal infection
XX Helicobacter polynucleotide and polypeptide sequences - useful to
PS Claim 1; Page 330-337; 362pp; English.
XX This claimed Helicobacter pylori polypeptide, designated GHPO 1484,
CC can be used in vaccination methods for preventing or treating
CC Helicobacter infection. 85 Helicobacter polypeptides (see
CC AAW1474-71558) are claimed, as well as isolated polynucleotides
CC (see AAV5209-93) that encode them. The invention also provides:
CC methods for producing these Helicobacter polypeptides in
CC recombinant host systems, and related expression cassettes, vectors
CC and transfected host cells; live vaccine vectors
CC that contain the polynucleotides of the invention and which can be
CC used to prevent or treat Helicobacter infection; therapeutic and/or
CC prophylactic methods involving administration of polynucleotide
CC molecules, polypeptides or monospecific antibodies; methods for
CC detecting the presence of Helicobacter in samples using e.g.
CC the polypeptides or monospecific antibodies; and methods for
CC purifying the polypeptides by antibody-based affinity
CC chromatography.
SQ Sequence 2893 AA:
Query Match 3.5%; Score 223; DB 19; Length 2893;
Best Local Similarity 20.5%; Pred. No. 0.00013;
Matches 281; Conservative 160; Mismatches 473; Indels 458; Gaps 73;
QY 14 GGLAVFG-----ALGSASFGKQSDKSNQTLVNOARTLDANSVRLAGLQNGSL 64
Db 287 GGYTFNKKEFSATNMTAFSSGFFNKGVSSFN-GTFSFNASYTD-NOATFONSSFNNGT 344
QY 65 F-----NIVLRVDFDNFTTANGTIKIDSTPKPLYGDLSDCGGYKVOIYSD- 114
Db 345 FTENNOTNPTNNAQHPOIIONSSFSGNATILK-----GFVNFQAEFNNS 387
QY 115 ---YTSRNFQDQRTAYALLVNDVANVHLKRTNSNRIGNNNNSKVIYGV----- 167
Db 388 NHQITQMSFNNAFTNNTGKITIEKDAFEN---NTFENFNSVDPNNNS--VTGGVTLGG 441
QY 168 -----DNAAHYIRFTDQGTKEFNFN-----OTQGEIVDFILDAPI--LP 205
Db 442 KNDLKGSTLDFGSSSKITTA-QGTTNLTSLGSEKSVTLNLSGGITLNLNHLNGLT 500
QY 206 KDLHPDWYNIQRIKILPNDVNTAVVWPVGRVSGTNADGMEFGNGQITNTPDIAQTK 265
Db 501 SAKLTN-----ESLSNQSFAQGLMDITTYNGV-----TGQLLN-ENAAATSK 541
QY 266 TTTDNQNPSTFNSGAMPANRRYDSOLNWKRI-KTSFOLDERINTNSRI-----GNRN 319
Db 542 PTDSPPSKSSTNSTQV-----YQGYKIGDTITIKLOETFSINSITIALMSGTYT 591
QY 320 -----NNSKFVIG-----VDNPAHYIR--FTDGTKEFNFTNQTQGEIYND--- 358

Db 592 PEPVINGSKFDLSASNTINADMWYDHKKYIPKQNFTEGTYLLPSVOIQMSYTNFSFKO 651
QY 359 -FIIDAPILPKDLHPDW--YVLYIQRIKLPDVNTAVV-----PWPYGRVSGTNADGME 410
Db 652 FFSANGSNLYIGYNSTWTDHNVSSGTVSFGDTGSLNMGCHGFWPYPYQCTGTT--NGTY 709
QY 411 DC-----GNGQITN-----TPPIAQRTYTTDNQNP-----TFNSGAMP 444
Db 710 SAYHYITANLRSGKRIGTGGAANIIFNGVDSINMANNTIQHAGAGIYSSSMFTSTQSDM 769
QY 445 GANN-----RYDSOLNWKRIKTSFOLDEREYVPEPTGSEENKNTIRLATGSLPSNERXYI 500
Db 770 NSQNLNGLNLSNGKLSAYGTTFTEKDKGFENGAQVFEFTN-----814
QY 501 LDIPGPTQYTLKEDSVNFSRLYLN-----SYNSLFTGDSIYIFGISLPSLM 549
Db 815 ---GGSYQ--FSGDLSNFSNNQFNSGSEISAKAKASFNMANFNNSAFNNSAFTSF 869
QY 550 YSFPTRLSDLTALNOVKTTDIEASSTGNTGTTTGTADTSSGSGTACGTNT----- 603
Db 870 VGDFTMANSNL-----QI-----AGNAVFGNSTGNTAMFNMTGSVNIISGNTFPDNYV 919
QY 604 ---TNTS-----QTVSNPT--LNTYRSFG-----IDSKP---TSA 630
Db 920 FNGPTNTSVKGYVTLNNTILKLNAPLSFGDGTTFNMAHSVINAESITNGNPIYLVSSS 979
QY 631 NKIDETWADPNYIARLYAEVRLQIENEIPITMAGNFRNTIGVGFTSGSRVYLKAS 690
Db 980 KEIPEYNAFESKNIQMLINTYQGH--GASSEKLVSSAGNGYDVV-----YS 1022
QY 691 YNGDQRTGPNFQPL-----YVEGLG-----YQQTRTG--TFWGYTK- 727
Db 1023 FNNQ--TYNEQVEFSONSISIRLGVNMFVQVDMKSDHLYQNALGFMTPNYSYNN 1079
QY 728 ---LNNSPY-----DVLDSPRVGTEPNOFRRTSLTPYMGGYLTBEARFESNTPYIR 778
Db 1080 NLGANNNTIYDKSIDFVYASGKTLFTKAFESOT-----FTGQSAIYFGAKSIWT 1130
QY 779 AOGDTPESRSIFGSGYDNT-----YE---YIO-----803
Db 1131 SLSDAPQSNFTIIRFG--DNKGAGSNDASGHCMNLQICGFTGHEAQKIYTGSIESGMR 1188
QY 804 ---SVLGFDDGIRN-----NLNVGVKASS--FLNSNPNRNGLEMTAATYLLRS 846
Db 1189 ISSGGGASLNFNGLOGLILTNATLYNRAAGTQSSNMNIS-----NSANTQONSIFYD 1242
QY 847 QIGLARTGDLRNOQPFQTHQVIVSPDQESSIKRTIFPGNOLWYFLFTNENKSS- 905
Db 1243 D--TQONGGNPN--FETNALNLDPSNPSFPGYGTQGSVFKFAKVAISFTNSTLSSG 1297
QY 906 VYTLR-----LADSSNPDASSFSPTSLIDVNEIGYI-----LPLDN 943
Db 1298 LYOMQAKSLFENSNMLSVSG--TSSIKAAVAILNSONASINASHSTLEQGLDNLVNDT 1354
QY 944 SFY-----TVNAGVAL--FSSNPGSGSYT--AVN-----TFNOMLSDAIEFGS- 985
Db 1355 SSLNLNOSTINVSNNATINDVASLASHSGSHLENFGAANFNSAMITTLNNSIYFGAV 1414
QY 986 --GAKY-----TSPDWTGTOFKPDEYLIQNGFSSQVARNFVYQOSFLNSLYDTPPANAG 1037
Db 1415 SLGGQTNLSNNSLSDQGS-----SATSNTAFNFYDN-AFSQSTIFTHQA--- 1459
QY 1038 TNYRVVYVDPDGNLTNQNLEPKVQIYLDQKKYDAKLKNNNVTFESYNNFAL 1089
Db 1460 LDIKAPLSIGNLNPN-----NSSYLD-----LKSQLY--FGDQGS 1496

RESULT 10
AAB46351
ID AAB46351 standard; protein; 2902 AA.
XX AAB46351;
AC AAB46351;

XX		05-APR-2001	(first entry)
DT			
XX		H. pylori HPN165 protein.	
DE			
XX		Microbial infection; antibacterial; Helicobacter pylori infection;	
KW		vaccine; screening.	
OS		Helicobacter pylori.	
XX			
MO		MO200073502-A2.	
PN			
PD		07-DEC-2000.	
XX			
PF		31-MAY-2000; 2000WO-EP05024.	
PR		31-MAY-1999; 99DE-1024965.	
XX		17-JUN-1999; 99DE-1027740.	
PR		21-JUL-1999; 99DE-1034029.	
XX			
PA		(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.	
XX		(CREA-) CREATOGEN GMBH.	
PI		Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;	
DR		WPI: 2001-049948/06.	
XX		N-PSDB: AAF25628.	
PT		Preparing an agent for diagnosis or control of microbial infection,	
PT		useful particularly against Helicobacter, based on identification of	
XX		essential genes in defective mutants -	
PS		Claim 37; Figure 15; 366pp; German.	
CC		This invention describes a novel preparation of an agent (A) for	
CC		detection, prevention and/or treatment of microbial infection by:	
CC		(i) identifying essential genes (I) and corresponding polypeptides	
CC		(II); (ii) identifying compounds that are directed against (II) and	
CC		inactivate the microbe; (iii) testing these for suitability for use; and	
CC		(iv) formulating selected (A). Identifying essential genes (I) comprises	
CC		preparation of gene-deficient microorganisms by conditional antisense	
CC		inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),	
CC		then determining viability and/or survival of the deficient organisms.	
CC		The products of the invention have antibacterial activity. (A) (which may	
CC		be a nucleic acid (ia), vector or host cell containing (ia), derived	
CC		polypeptide (Iia), or fragments, (Iia)-specific antibodies or their	
CC		fragments or an inhibitor of (Iia)) are particularly used for diagnosis,	
CC		treatment or prevention of infection by Helicobacter pylori. Particularly	
CC		(Ia) and (Iia) are used in DNA, subunit or live vaccines. The method	
CC		identifies essential genes, including those that have homologs in other	
CC		species, so identified (A) should have a broad spectrum of activity. Many	
CC		gene-deficient cells can be screened quickly, in an automated process,	
CC		and the identified genes can be used for screening without purification.	
XX			
Sequence		2902 AA;	
Query Match		3.3%; Score 214; DB 22; Length 2902;	
Best Local Similarity		19.4%; Pred. No. 0.00046;	
Matches 277; Conservative 157; Mismatches 449; Indels 542; Gaps 70;			
OY		26 SFGKQSKNSKDNDQLVQAR--ILDANSVRLAGGNGSLFNTYLKVVDNFITANGCT 83	
Dd		1 : : : : : : : : : : : : : : : : : : : : : : : : :	
OY		260 SYSKGKTNNAT-NTFSSNSSGFPEENAT-FSGAKLNGGAFTF-----NKKFNATNMT 311	
Dd		84 IIKLDSF----TKPLYGLDLSDDCGGYVKQIVSYTTSRRNFDROTRAVYALLVNDEA 139	
OY		1 : : : : : : : : : : : : : : : : : : : : : : : : :	
Dd		312 AFNSGSFFPKGTSSFNGANFSN-----ASYTFNNQATFQNSSFFNGCTFTFPNDQT 360	
OY		140 N-----VHLKRITNSNRIGNRN----- 157	
Dd		361 NSQTHPOLIONSSPFGSATTLKGFAITFEQAFNNSHOLTIONASFNNAFTNNTGKITTEK 420	
OY		158 ----NNSKF-----VIQGV-----DNPAHVIRFTDDGTKEFNTN-- 187	

Db	421	DASENNITSEFNPVDTNNMTISGVTYLSGKNDLKNGATILDPSSKTIITL--QGTTEWISLG	479
Qy	188	-----QTQGEIYNDFILDAPILPKLDHPWNYLYIQRKLPNDVT-----A	229
Db	480	SEKSVTILNSGGIITYNHLNHLAT-----NSLTKALTESSSKPSQSA	523
Qy	230	VWPEPVGVSCTNADDCMFCGNGOINTPDLAOTKTTTNDONPSTENSGAMPGANRRYD	289
Db	524	QGLWDMITLYNCV-----TGQLLN--ENAAATSKPTDSSPSKSTSTVOY-----	564
Qy	290	SQWVWKHRI--KTSFQDLDERIMTNSRI-----GNNN-----NNSFEVYIGVD-----	330
Db	565	--YOYGYKIDPTIKYKLOETFSHNSIITQALESGIYTPPPVINGSKFEDLSASNYINADMPW	622
Qy	331	-NPAHVI-----RFTDDGTFKFNFTNTOGEIYNDFILDAPILPKLDHPWNYLYIQRKLP	385
Db	623	YNHXYIYPKSNLFESGTYLPSVQWMSYTNSEFQTSASNSNLVIGYNTATWTDHANNVS	682
Qy	386	ND-----VNTAVPWPVGVGVSCTNADDCMFC-----GNGQITN	419
Db	683	SDTVAFGDTSALNGHCGPWPYXOCTGT--NGTYSAYHYITANLRSGRNIGTGGAAN	740
Qy	420	-----TQDIAQTKTTNDONPSTENSGAMPANNRYDQO-----LWVKRRIK-----TSF	464
Db	741	LIENGVDISINATNATTOHNAAGVSSSKTFSTQONNDONQNLNGLNSGRKLTYGTTFTNQ	800
Qy	465	QLDEKFVPEWTGSEENKNTIRLATGSLPSNERWYIIDPCTPOVTLKEDSVNYSRLYL	524
Db	801	AKDKGFIENAOATFENNTFN-----GGSYO--FSGDLSLNFENN--	837
Qy	525	NSVNSLSF--ICDSIYIFGTSLSPLMWYSF-----PRLSDLTALN--QVKTDDIAS	574
Db	838	NOFNSGSFIEIAKNTIENNANFNSTSNFNSATTSFVDFPTNANSNLDI-----AG	891
Qy	575	STDNGTNGTNGTTTAD--TSSGSTG--AGT-----GWTJNTNS--QTVSNPT--LNT	617
Db	892	NAVGNSTNGSONTANFENNTGOSVNIAGATENDVYFNSPTNTSVGAKVTLNNITLKNLNA	951
Qy	618	YRSFGIDSKPTSANKIDE-----TNMADPNYIARIVAEYR-----LGIQ	657
Db	952	PLSRQDGTIVPSAHSVINIGEAITNGNITLYVSSSKALEYNDAFSKNLMOIINYGHGAS	1011
Qy	658	NEIPTTANGN-----FIRNTIG--GYGF-----	678
Db	1012	SEKLVSSANGCVYDVYFSPNNOYTNFOEYFSPNSISIRLGVGMFEDYVDEKSDRLYQ	1071
Qy	679	TSTGSRVYLRASYNGDQRPCTGNFOFLY-----VFCYLGVOORTG--TF	721
Db	1072	NALGPMYTPMSYNN--LGLNLTITYYDNSIDFYASGKTLFTKAEFSOTFTQONSAT	1122
Qy	722	WYGYTKLLNNSPYDVLDSPR-----VGTENQD--RRT	752
Db	1129	VFGAANNITMS-----VSDAPQSVIIRPBDNKGAGSNDASGCHWMNQIGFTIGHAEQKI	1184
Qy	753	SLTYPVMGGYLTBGCARFSN-----TPYIAQDTPPSRSIIPQSYSDNTE	800
Db	801	YIQSVLGGDGIIRNNLNVGVKASSFLN--SNRPN--NGLNEMIAATTYLRSQIGLART	853
Qy	1234	-----VSNSANIDAQNSYFIDDTAQNKGNCNFSFNALNIDFSNMSFPGYVG--	1275
Db	854	SGLPNQOFGFTHQIVISVPGQFSIKNIRTIPEGNOLWYLFITFNEKKKS--VYTLR--	910
Qy	1280	-----QTOGVFKFNAVNAISFTNNSNLSGLYQMAK	1311
Db	911	--LADSNPDASSFSPTSLIDVNEIGYI-----LPLDNSFY--	946
Qy	1312	SVLPFNSMLSYSG--TSSIKAMININSONASINASHSTLELQGLDNLNDTSSLINMQ	1366
Qy	947	-TVNAGVAL--FSSNPGSPQSYT--AVNTEFQNL-----SDIAFEGS-----	985

Db 1369 SAINVSNATINDYASLASNGSHLNFNGAVNENSANITTSISSSIVEKVASLGQEN 1428
 QY 986 -GAKYTSDFMGTIQFKPDEYLLIONGFTSOVARNEFTNOSFLNVDFTPANAGTNRVY 1044
 Db 1429 LSNNSLDFOQS-----SATSNTAFNFDN-AFSQSITPFOA---LDIKVPL 1473
 QY 1045 DPGNLTNONLEPKAYOIYLDGKKYDAKLRNNNLVTSYNNFAL 1089
 Db 1474 SLGMLNLPNNSSVYL-----NLKNSQVLF--FSDQSL 1503

RESULT 11
 ABB54070
 ID ABB54070 standard; protein; 1180 AA.
 AC ABB54070;
 XX 16-MAY-2002 (first entry)
 DT Lactococcus lactis protein yhpD.
 DE Lactococcus lactis protein yhpD.
 KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX Lactococcus lactis I11403.
 OS
 XX FR2807446-A1.
 PN
 XX 12-OCT-2001.
 PD
 XX 11-APR-2000; 2000FR-0004630.
 PF
 XX 11-APR-2000; 2000FR-0004630.
 PR
 XX 11-APR-2000; 2000FR-0004630.
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX Bojoltine A, Sorokine A, Renault P, Ehrlich SD;
 PI WPI; 2002-043418/06.
 DR
 XX New nucleotide sequence useful in the identification of Lactococcus
 PT lactic and related species -
 PS
 XX Claim 6; SEQ ID No 772; 2504bp; French.
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB5621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 CC
 XX
 SO Sequence 1180 AA;
 Query Match 3.2%; Score 207.5; DB 23; Length 1180;
 Best Local Similarity 21.0%; Pred No. 0.00034;
 Matches 246; Conservative 132; Mismatches 445; Indels 347; Gaps 59;

QY 15 GLAVFGALGSA-----SEGFQSKDSKNDNTOLVNOARFLDANSVRLA----- 56
 Db 118 GLGIFSVGSATSATGFLKSTIKYDSSGSSSLDPTKSGDAIGVLTGASKAQAALANAOS 177
 QY 57 -----GLGONGSLFTVTL--RDVDDNFTTANGTILIKDSPTKFLYGLDLSDDCGGYKV 108
 Db 178 SQATNDGLGIRG-LPPTVFLGRDLYSNLSNVSTGS-----LTGIDSATVNGW-- 224
 QY 109 KQIVSDYTSRNFQDQROTFRAYYALLVNDEANVHLKRINTNSNRIGRNNNSKEYIGVD 168
 Db 225 -----TDPGVVMAIRMTDSSGTLQAANPSSL--GSA 254

QY 169 NPAHVIRETDGCTFENFTNOTGSEIVNDPILDAPILPKDLHP--DWMYLYIQKILPND 225
 Db 255 AYPSPVNTONGATYATATDTSYALG---VASITTSVOEPLTW-----TPDA 301
 QY 226 VNFVAVPPVPCRVSGTNAADDGMPDCNGOITNDPILAQITTTTNDONPSTFNSGAMPAN 285
 Db 302 TMTA-----STGYTSOT-----LTFMLIAQT-----TSGALIGASAVNGCT 338
 QY 286 NRXY--DSOLNKKRHKIKTSFQDLDERINTNSNRIGRNNNSKFVIGGVNDNPAHVIRETDDGT 343
 Db 339 GGYTLTHTHNLONSILVGFV-----GGTGGVNGLSVSLNGA-----TITGS 380
 QY 344 K-----ENFTNOTGSE-----IYNDFILDAPILPKDLHPDMYNTLYIQKILP 385
 Db 381 KCTEPVPIYTLNKKYGGQPIPSMAOSSITANVNDTIGALTLAPSSSTPD----- 428
 QY 386 NDVNTAVVWPVPRVSGTNAADDGMPDCNGOITNDPILAQITTTTNDONPSTFNSGAMP 445
 Db 429 NMTYTVVVP---AAPGYQANGA-----TTTTVLAAMPASTTNTSATTV--SNFTVG 475
 QY 446 ANNRYSOLNKKRHKIKTSFQDLDERINTNSNRIGRNNNSKFVIGGVNDNPAHVIRETDDGT 505
 Db 476 TTN--PNAINV-----STPSQOYVLFEMAKASG---TTLA--NLPNQIAYGATGT 521
 QY 506 TPQVTLKEDSVNVFSRLYNSVNSLSFICDSIYITGTSELPSLWYSPFRLSDLTALNQ 565
 Db 522 ASVVT-----DNFNSNTLLS-----STATALS-----NA 546
 QY 566 VKTDDIEASSTGDTTNGTTTADTSSGAGTNTNTSQTV--SNPTLNTYRSFGID 624
 Db 547 VPT-GYNTITQISNGTITVSGATTAALTA-----FTSQPVTGATSTLNTV-- 591
 QY 625 SKPTSAKKIDETWADPNVIEARIYAEYRLGIONEPIPTNAENFIRNTIGGVTGSGR 684
 Db 592 --TWTAAVSQT-----ATYVGYTSTALNTPALPALTISG-----TTGS 631
 QY 685 VYLRASYN-GDORPTGNFOPFLVFGYLGQOYRTGTFWYGYKLLNSPYVLDSPRYG 743
 Db 632 VTPTGTNTSPQIPSGYISAIY-----ACTSASGT-----LISSSTT 669
 QY 744 TETNOFRRTSLTYVPMGVYLTREGARSEFTPIRAQDTPES--RSIFQSGYSDWYBY 801
 Db 670 NGTNAAMPSCGVYAPTNQYIQLAPMTVAVTWFTISIDPNDDPLNAVAGGFTTP-- 725
 QY 802 IQSVLGPETGKNNLNVGKASSFLNSRPNNGHEMIATYTLRSQIGLAR--TSGLPNQ 860
 Db 726 VGSILG-----GSTTNYQSYVDNMSLSLGNLTSTDAVNTYDGAITGYOYOE 780
 QY 861 PEGTTHQVIVSPGDDFESSIKNIRTIFFPQOMLWFLFTNNNNSS--VYTLRLAD---SS 915
 Db 781 --GTTTATNAPATITTYKTFESDLAHP-----YVLAGTSSNRQLVHLMYVDOTTTIS 933
 QY 916 NPDASSSP--TSLIDVNEIGVILLPLDNSFYTVNAGANVALFSSNPGSPSYAVANTEN 974
 Db 834 NSTSQOHVNPATNYPSSSEL-----TSTSDIDGADDDGL-----SY-AINGVY 875
 QY 975 ONLSDIAFEGSGAKYTSDFMGTIQFKPDEYLLIONGFTSOVARNEFTNOSFLNVDFTPA 1034
 Db 876 ---AIVNITGGGTGWDGATNSLQISICTYT-----ENYVALNLTGVAATNML--PS 923
 QY 1035 NMG---TNRVYVDDGNTNQNLPKAYOIYLDGKKYDAKLRNNNLV--TFSYNNFGA 1088
 Db 924 HSGGVISOPLQTLDAF--SLANDTVISQTTLEVQDA---TALTNNHHIIPRAASYNPSSD 979
 QY 1089 LPSWVPTAIGSTL-GILAIMILIGAI 1117
 Db 980 LVG--ASNADGSSVENIETNPVINGDAIGV 1007

RESULT 12
 AB008784
 ID AB008784 standard; protein; 2122 AA.

XX	AB08784;
AC	
XX	28-MAY-2003 (first entry)
DT	
XX	Moraxella catarrhalis outer membrane protein, OMP106.
DE	
XX	Outer membrane protein-106; OMP106; Moraxella catarrhalis infection;
KW	vaccine.
XX	
OS	Moraxella catarrhalis.
XX	
FH	Key
FT	Location/Qualifiers
FT	Peptide
FT	Protein
FT	Region
FT	Misc-difference
FT	/note= "Specifically claimed in claim 9"
FT	/note= "Encoded by ATCTAGGC"
PN	US2002177200-A1.
PD	
PX	28-NOV-2002.
PF	20-MAR-2001; 2001US-0813214.
PR	12-NOV-1997; 97US-0968685.
PR	03-MAY-1996; 96US-0642712.
PA	(ANTE-) ANTEX BIOLOGICS INC.
PI	Tucker K, Plosilla L;
DR	WPI: 2003-328486/31.
N-PSDB:	ABX93525.
PT	Novel outer membrane polypeptide of Moraxella catarrhalis, useful for
PT	producing an immune response in an animal, and as ligands to detect
PT	antibodies elicited in response to Moraxella infections -
PS	Claim 33; Page 28-34; 51pp; English.
XX	
CC	The invention relates to an isolated or substantially pure outer membrane
CC	polypeptide of Moraxella catarrhalis (OMP106 polypeptide). The
CC	polypeptide and its peptide fragments are useful for producing an immune
CC	response in an animal, and as ligands to detect antibodies elicited in
CC	response to M. catarrhalis infections. The polypeptide and its peptide
CC	fragment are further useful as active ingredients in vaccines against M.
CC	catarrhalis infections. The polypeptide is useful to prepare antibodies.
CC	An antibody binding the polypeptide is useful in immunoassays to detect
CC	M. catarrhalis in biological specimens, and also in passive immunisations
CC	against M. catarrhalis infections. An antibody binding the polypeptide is
CC	also useful to facilitate isolation and purification of the polypeptide
CC	and its peptide fragment, and as probes for identifying clones in
CC	expression libraries that have inserts encoding the polypeptide and
CC	peptide fragment. An antibody binding the polypeptide is also useful to
CC	diagnose M. catarrhalis infections. The polynucleotide encoding the
CC	polypeptide is useful as a probe to identify the presence of M.
CC	catarrhalis in biological specimens by hybridisation or PCR amplification
CC	and also to detect other bacteria that might encode a polypeptide related
CC	to M. catarrhalis OMP106. The polypeptide, an antibody binding the
CC	polypeptide, its peptide fragment and the polynucleotide encoding the
CC	polypeptide are useful as reagents for clinical medical diagnosis of M.
CC	catarrhalis infections and for scientific research on the properties of
CC	pathogenicity, virulence and infectivity of M. catarrhalis, as well as
CC	host defence mechanisms. The present sequence represents the amino acid
CC	sequence of M.catarrhalis outer membrane protein (OMP)-106.
XX	
XQ	Sequence 2122 AA;

Query Match	3.2%	Score 205.5:	DB 24:	Length 2122:
Best Local Similarity	19.3%	Pred. No. 0.001:		
Matches	282:	Conservative 156:	Mismatches 502:	Indels 507: Gaps 70:
OY	24	SASFGFKSDSDSNTOLTQVNOARTL-----DANSRLAGLQGN-----GSLFNTVLRD	71	
DB	482	AATYG--QKKVYNQTAESALQTFYKKYKDKNGNDANDSKITTVGKNNKPPDQYNT-----	535	
OY	72	VDDNFETPAANGFTIILKDSFTKPLVGLDLSDDCGGKVKVQIYSDYTSNRREDORATRAY	131	
DB	536	-----LKL-----KGGNGVDVTTETNG-----PYTF	556	
OY	132	ALLVNDGANHLKRIKNTNSNRIGNRNNNSKVIIG-----GVDNPAHVR	175	
DB	557	GLANONGLTJVYNGSTLNDGLSVKNTNSKKQIQVAGDGTTFYDINSKPGAGIEN-----TTR	613	
OY	176	FTDDGTRKNFNOGGEIVNDFILDAPILPDLHPDWNLVYQRIILPNDVNTAVVPAPV	235	
DB	614	ITRPG--IGFANNNGS-----LDAN-----KPLITPTGIN-----A	642	
OY	236	GRVSGTNADDMFDCGNG-----QITNDPIAQKTTTDD-----NONPSTENGAMP	282	
DB	643	GKKELTNNQSAINPATNGQLDPMNRSLSTAMEKSGSAATIKDLNLNSQVLETFAGDGRP	702	
OY	283	GANNRY-----DSQLNVKHRITSFOLDERITNSNR	314	
DB	703	NVTKKLEILKVKGGKATTADTLKNNIGVADSTDSNLSLTV--LAKTILSDLD--AVNTKTLT	760	
OY	315	IGNR-----NNNSKFFVIG-----GVDNPAHVIRFPD-----DG	342	
DB	761	ASDKYTVYDSGNNTAKLQNGDLTFESKONTGATPATNSKITGVG-----LKFDTNNGIALDG	816	
OY	343	TKFNFTNOGGEIVNDFILD--APILPKDLHPDWNLVI--QRKILPNDVT--AVVEWP	396	
DB	817	TTY--ITKQKVFAPQDQSLDKSKPFLDQ-----KLKVGGEIITNIGMNGKAITGLS	869	
OY	397	VGRVSGTNADG-----MFCGNGQITNTD--PIAQKTTTDDQ	433	
DB	870	NLTLDATNATGHTQTLQGLVDSTDKTRAASIGDVLNAGFNLKNNNDARKDEVSTYDTVD--	927	
OY	434	NPSTFNSGAMPGANNRVDSOLNVKHKRITSPQLDKEFYVPEWTSSEKKNITRLATGSLP	493	
DB	928	---FINGNATFAKVTYDGKAS---KVAIDVANDGTTIHLTGADGKNQNGIYATTTLLT	978	
OY	494	SNERYMILIDIGTFQVYTLK---EDSVAVFSRLVLSVNSLSFIDSIYIFGTSPLSLM	549	
DB	979	KT-----DAGDKRAINSVNSGDCKALKLNKDIADNLMTLA--GEINTNGTAD-----	1022	
OY	550	YSPFTRLSDLTALNVKQTTDDIEASSTJNGTGTNGTTTADTSSGSGTAGTCNTNTSQT	609	
DB	1026	-----TALQTFQVKRVK---ENGDDDN---DADITV-----VGKDAKTQD--	1055	
OY	610	VSNPTLNTYRSFGIDSKPTSAKIDETMMADPNVIEARIYAEYRGIONEIPITNAGNFI	669	
DB	1060	-----VNTLKLKGGKGLDIQTK-----DGVTTFGINQSGAKAGNMTTLNNGLSI	1106	
OY	670	RNTIG-----GVCFSTGSRVYLRASYNGDQRPNGFQEPFLYVFGYLGQOQTRPGT	720	
DB	1107	KNTAGNEBQIQVGAQDQVAKYVANGVY--GAGIDGTRITRDELGFAGCTNGSLDKSKPHLSK	1166	
OY	721	--FWYGYTKLNNSPYDLVDSFRVGETTNOQFRTISLTYPVMGYV-----TE-----	768	
DB	1166	DGINAGKAKTN-----IQSGEIAQNSND-----AVNGGKYDLTKLELNKISSTA	1211	
OY	769	RFSFNTPIIRAGQGPBPESRSIFQSGISDYNTYEXIOSYLGCF--DQIRNNLN-----	820	
DB	1212	KTAQNSLHEFVADQGNFTVSNPYSSYDTSKTSYDVTTFAGENSIITTKVNGVAVGID	1271	
OY	821	ASSFLNSNR-----PNNGLEMTAATYLRSGIARTSGLPNQOQPGTTHOVISVPGDOF	877	
DB	1272	QTKGLTTLTKLVGNNNGKGIY-----IDSQNGQNTITBLSN-----	1307	
OY	878	SSIKNIRTIPEGNQIMVFLFTENNKKSSVYTLR---LADSSNPASSFSPTSLIDVNE	933	

```

DB      1308  -TLANV-----TNDKGSVRTTEGKTIKEDDKTRAA-----SYDVLS 1344
OY      934  IGVIL-----PLID--NSRYTVNAAAGNALFSSNGSGSYTAVTFEONLSDIAFEGSGA 987
DB      1345  AGFNLOGNEAVDEFTVDVNFADGNATTKATYTDTSKTSKYVYVNVDDTTIEVKDK 1404
OY      988  K-----YTSDFMGTIOFKPDEXLIQNGFT-----SOVARNFVYNOSFLNSVDFTP 1033
DB      1405  KIGVTKTTLTSTGTGA-----NKFALSNQATGDALVYKASDIVAHNLNTLSGDITQAKGASQ 1459
OY      1034  ANAGNTRYRVVPPDGNLTNQNPLTKVQIQYIDGKYDAK-----LKNNLVTFEY 1083
DB      1460  ANSSAGY---VDADG-----KYVDSIDNKYQAKNGDYDKTKREYAKDKLVAAQ 1508
OY      1084  NNGALPSWVVPYTAIGS-----NLGLTAL 1107
DB      1509  TPDGTILAKDNLNLSVNAAGTKRIDKGVSFVDSGQAKAMPVLSANGLDLGKVISNV 1568
OY      1108  MTLGLALGILPRAQKLO---DKGKTFEKKV-DTLTAAVGSYKKIITO----- 1154
DB      1569  V-----AQTPLTPAGDTG--TTAKKLGELTITKGGQDTNKNLTNNIGVAGT 1614
OY      1155  ---TANYKKRPAALGAKSG---DKRPAAAKPAAPAKSPAKSPAKPTGPK-----S 1203
DB      1615  DGFYVLAKDLTMLNSVNAAGTKRIDKGVSFVDSGQAKAMPVLSANGLDLGKVISNV 1674
OY      1204  GAPPKPT 1210
DB      1675  GKTKKPT 1681

RESULT 13
AAU37120
ID      AAU37120 standard; Protein: 2344 AA.
XX      AAU37120:
XX
XX      14-FEB-2002 (first entry)
XX
XX      Staphylococcus aureus cellular proliferation protein #1290.
XX
XX      Antisense: prokaryotic cellular proliferation protein;
XX      antibiotic; antibacterial; drug design.
XX
XX      Staphylococcus aureus.
XX
XX      WO200170955-A2.
XX
XX      27-SEP-2001.
XX
XX      21-MAR-2001: 2001WO-US09180.
XX
XX      21-MAR-2000: 2000US-191078P.
XX      23-MAY-2000: 2000US-206848P.
XX      26-MAY-2000: 2000US-207727P.
XX      23-OCT-2000: 2000US-242578P.
XX      27-NOV-2000: 2000US-253625P.
XX      22-DEC-2000: 2000US-257931P.
XX      16-FEB-2001: 2001US-269308P.
XX
XX      (ELIT-) ELITRA PHARM INC.
XX
XX      Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX      Yamamoto RT, Xu HH;
XX      WPI: 2001-611495/70.
XX      DR      N-PSDB: AAS54979.
XX
XX      New polynucleotides for the identification and development of
XX      antibiotics, comprise sequences of antisense nucleic acids -
XX
XX      Example 3; Seq ID No 12713; 511pp; English.

```

```

XX      The invention relates to antisense inhibitors of genes essential to
CC      prokaryotic cellular proliferation, their use in identifying the
CC      genes, their use in the discovery of novel antibiotics, the essential
CC      genes themselves and the encoded proteins. The prokaryotes used are
CC      Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC      pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC      invention is also useful for the identification of potential new targets
CC      for antibiotic development. The antisense nucleic acids can also be used
CC      to identify proteins used in proliferation, to express these proteins,
CC      and to obtain antibodies capable of binding to the expressed proteins.
CC      The proteins can be used to screen compounds in rational drug discovery
CC      for homologous nucleic acids which are required for cell proliferation in
CC      essential prokaryotic cellular proliferation protein.
CC      Note: The sequence data for this patent did not form part
CC      of the printed specification, but was obtained in electronic
CC      format directly from WIPO at
CC      ftp.wipo.int/pub/published_pcl_sequences.
XX
XX      Sequence 2344 AA:
XX
XX      Query Match 3.2%; Score 205.5; DB 22; Length 2344;
XX      Best Local Similarity 19.2%; Pred. No. 0.0012;
XX      Matches 226; Conservative 167; Mismatches 483; Indels 301; Gaps 50;
OY      2 NISKKIKSYTL-----IGGLAVFAGLGSASFQKOSDK-----SNDNTOLVN 43
DB      58 SISKKMGYGLKTTAVIGAFVYVNLMDQO-AFAASDAPLTSSELTOSEYVGNOSTTID 116
OY      44 -QARTLDANSVRLAGLQNGSLFNTVLRVDNDFITANGTITIKLDSFTKPLYLGLDSD 102
DB      117 ASTSTADSTSVTKNSSSVOTSNSDIYVSEKSEVISTSTNSQOEKLTS----- 166
OY      103 CGGYKVKQIAYSDYTSNRREDQROTAYVALLVNDPANHLKRTNNSRIGNRNNNSKF 162
DB      167 -----TSESTSKNTSSSDTKSVSTSTSEDP-----INTSTQSTASNNSTOS 211
OY      163 VIGVDNPAHYIRTFDDGTGFNFNTOTGIEVNDFLIDAPILPKDLHPWNYLYQKIL 222
DB      212 T-----TPPSANLKTSTTSTS-----TAVYKLRTPS-----RLAMSTF 245
OY      223 PNDVNTAVPWPVGRVSGTNAADGMFDCGNGOI-TNTDPLAQKTTTNDONSTFNSGAM 281
DB      246 ASATTTALNTANTITVKNKDLKOYMTSGNATYDOSGVVTLTQDTTSOKGAILTGT 302
OY      282 PGANNRYDSQLNVKHKRITKTSFQDLERINTNSRIGNRNNNSKFEVIGVDNPAHYIRFTD 341
DB      303 -----RIDS-----NKSFFHSGKYNLGNKKGEGHGN-----GG-----D 330
OY      342 GTFNFNTOTGGE-----YNDFLIDAPILPKDLHPWNYLYQKILPNDVN 389
DB      331 GIGFAFSPVGLGETLNGAAGVIGLISNAFPKLDYHNTSTP---NSSKAKADPSNV- 386
OY      390 TAVVWPVGRVSGTNAADGMFDCGNGOITNTDP--IAQTKTTTND-ONPSTFNSGAMPGA 446
DB      387 -----AGGGAf-----GAVTTDSIGVASTVSSSTADNAAKLN--VOPTN 425
OY      447 NNRYSQNLVKKHKRITKTSFQDLERINTNSRIGNRNNNSKFEVIGVDNPAHYIRFTD 484
DB      426 NTFQDFDIN-----YNGDTRKVMYVYAGGTWRNLSIDIAKSGTTNLSLSTASTGCG 477
OY      485 -TRLA-----TGSLSNBERMYLID-PGTPQVTLKEDSVNVFSLYL-NSVNSLSF 532
DB      478 ATMLQOVQGTGTEFEYTESAVTVQVY--VDYTTKDLIDIPKRTYSGNVDDVYVTDINOQSALTA 535
OY      533 IG-----DSIY-----IFGTSELPSLWYSFPTRLSDLTALNOV-----KT-DDI 571
DB      536 KGYNTSYVDSVASTYNDNKKYKMNAGOSYTYFTDVKAPTVVAGQTLVEGKTMNPI 595
OY      572 EASSTDNGTTTNGTTTATDTSSGSGTAGGNTNTNSQVSNSTLNTMYRSFGIDSKPTSAN 631

```

```
Db 596 VLTTDNGTGVNTNTVTLGSLSDASANSIIIGPTKIGQSTVTVSTDOANKSTTTF 655
Qy 632 KIDETNMADPNVI-----EARIAEYRLQIOMEIPIT-----NAGNFIRNTI----G 674
Db 656 TINVDVTPAPYTPPIIDKSEVFS-----PLSPINIAIQDNSGNAVNTTVGLDS 705
Qy 675 GVGFTSTGSRVYLRASVNGDORFPGNFOPFLYVGYGQQRTGTFPMYGYTKLNNSPY 734
Db 706 GLTFDSITNMTI-----SGPTNIGSTITITVSTDASGNKRTTTF---KYEVTNNSMS 754
Qy 735 DVLDSFRVGTETNOFRT-----SLTYVMGCIYTEEGARFSNTPYRAQCDPEBSN 788
Db 755 DSV-STSGSTQOSQSVSTSKADSQSQASSTSGIMTSTA-STSKSTSVSLSDSVASKS 812
Qy 789 IFOSGSDMTYEXIGSVLGFEDGIRNLNLYGVKASSFLNSRPNGL-----M 837
Db 813 LSTSESNVSSTSTSLVNSQSVSSMSGSKSTSLDPIFINSSTESSESVSTSDS 872
Qy 838 IATYLRSGIQLARTSGLPNOQPGT-----THQVISVSPGDPFSSIKNI----- 883
Db 873 LRSTSLDSVSMSTSGSLSKSQSLSTSDASSTQSVSDSTNSISTSESLSESGSTS 932
Qy 884 RTTFPGNQLMYFLTFENKNSVYTLRLADSNPDASSSEPSLIDVNEIGVILPLDN 943
Db 933 ESISISNSI-----SNSVASSTSKL-ESQSTSLSTSDSKSMSTSE-----SLSDS 978
Qy 944 SFYTVAGNVALFSSNPGSPGYTAVNTFNOMLSDIAFEGSGAKTSDPWGTFIQKPD 1003
Db 979 TSTSDSVSGSLSVAGSQSVSTSDSMST-SEMISD-SMSTSGSLASDS-KMSVSSSM 1035
Qy 1004 YLIQNGFTSQ-VARNFVTNQSFNLSDVFTPANAGTN 1039
Db 1036 STSGSGSTSESLSDSISTSDSKSLSTSGSGSTS 1072

RESULT 14
ABJ19119
ID ABJ19119 standard; Protein; 10498 AA.
XX
AC ABJ19119;
XX
DT 06-MAR-2003 (first entry)
XX
DE Pathogen specific antigen related staphylococcal protein SEQ ID No 440.
XX
KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
KW autoimmune disease; HIV; hepatitis.
XX
OS Staphylococcus sp.
XX
PN MO200259148-A2.
XX
PD 01-AUG-2002.
XX
PF 21-JAN-2002; 2002MO-EP00546.
XX
PR 26-JAN-2001; 2001AT-0000130.
XX
PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX
PI Weinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;
PI Minn DB, Vytvytska O, Etz H, Dryla A, Weichart T, Hainer M;
PI Tempelwater B;
XX
DR MPI; 2003-075410/07.
XX
PT Identifying, isolating and producing hyperimmune serum-reactive
PT antigens from a pathogen, for preparing vaccine or medicament for
PT treating or preventing e.g. staphylococcal infections, comprises
PT providing antibody preparation
XX
```

```
PS Example 7; Page 234-235; 252pp; English.
XX
CC The invention relates to a novel method for identifying, isolating and
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC allergen, a tissue or host prone to auto-immunity, where the antigens
CC are used in a vaccine, comprises providing antibody preparation from a
CC plasma pool of a type of animal, or individual sera with antibodies
CC against the specific pathogen, tumour, allergen, tissue or host prone to
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC the 62 sequences of 53-2261 amino acids fully defined in the
CC specification, or their hyperimmune fragments are useful for the
CC manufacture of a pharmaceutical preparation, particularly a vaccine
CC against staphylococcal infections or colonisation against S. aureus or S.
CC epidermidis. The preparation of antibodies is useful for the manufacture
CC of a medicament for treating or preventing staphylococcal infections or
CC colonisation against S. aureus or S. epidermidis. The antibody
CC preparations may also be used for diagnostic and imaging purposes. Other
CC conditions that can be treated include cancer, autoimmune diseases or
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or
CC protozoan pathogens. This sequence represents a staphylococcal protein
CC relating to the method for identifying and producing pathogen specific
CC antigens of the invention.
SQ Sequence 10498 AA;
Query Match 3.2%; Score 205.5; DB 24; Length 10498;
Best Local Similarity 19.3%; Pred. No. 0.0098;
Matches 263; Conservative 152; Mismatches 507; Indels 441; Gaps 65;
Qy 40 QLVNQAARTLDANSVRLAGLQNGSLFNTVLRVDVDFNTAANGTITIKLDSFTK----- 92
Db 1104 QTVNAKATVATVIAK-----GQSFS-----IGDIKYGF-TLSNGQPIPSGFMTSDRTI 1153
Qy 93 -----PLYGLDSD-----DCGYYKKQIYSDYTTSRNPFDOQRATAYVALY 135
Db 1154 PTAQEVSOAMNAGTQLYHITATNAYHKDSEDFYLSLTIID--VKQPEGDQVRYTSRYDLP 1211
Qy 136 NDEANVHLKRIINTNSRIGRNNNSKFVIGGVNPNPAVIRFTDDGTFNFTNOGELVN 195
Db 1212 TDEIS-KVKAQAFINARBDVTTLAEGDISVYNTFNGANVSTIYNINKGRITKFSANLAN 1270
Qy 196 DFILDAPIPLPKDLHPWYNLYIQRK-----LIPNDVNT- 228
Db 1271 MNFLRWVNPFDYTVVWTNAKIANRPDGLSMSDBHKSLIYRYDAFLGTQITNDILTM 1330
Qy 229 --AVVFWPVRGVSQTADNDDGMPDCGNGQITNTPIAQTKTTTNOANPSTNSGAMP----- 282
Db 1331 LKATTTVPGLRNMITGNEKSOAEGRPNFRRTGYSSNATPDGQROFTLNGGVIOYLDI 1390
Qy 283 -GANNRYDSQLNVKHKRIKTSFOLDERINTNSNRIIGNNNNSKFVIGGVNPA----- 333
Db 1391 INPSNGYGGQ-----PVYNSN--TRANHSNSTVYVNVNPAANGAGAF 1431
Qy 334 --HVIREFTDGTKEFNTQGEIVNDFILADPLPKDL-----HPDWYNLYIQR 381
Db 1432 IDHVY-----KSNSTHNASDAVYKAQLVLTVPYGRPROVEHILNONTGNTDAINITY--- 1481
Qy 382 KIILPND-VNTAVVPMVPGRVSGTNADDGMPDCGNGQITNTPIAQTKTTTND--QNPST 437
Db 1482 -FVPSDLVNPTI--SVGNVT-----NHQVSGETFTNTITANDNGVQSVTV 1525
Qy 438 FNSGAMPANNRYDSQLNVKHKRIKTSFOLDERFVYPEPMTGSEENKNTTILATGSLPSNER 497
Db 1526 PNTSQTITGVDN-----NHQHVATA-----PNTV-SANKTKTINLAT----- 1562
Qy 498 YWILDIPTGPQVTLKEDSVNVFSRLYLNSVNSLSFGDSIYIFGTSPLSLMYSEFRL 557
Db 1563 ----DTSGNATAT--SFNV-----TYKPLR--DKYRVGTSS----- 1590
Qy 558 SDLTALNQVKTDDIEAST--DNGTTTGTTTADTSSGSGAGAGNTNTSQYNSNP 613
Db 1591 --TAANPVRIANISNATVSOADQTTIINSLTFTETVPNRSVARASANEI-TSKIVSN- 1645
```

QY 614 TLNTRSGIDSKPTSAKIDETNMADPNVTEARIVAEVRLGIONEIPITNACNTRNTI 673
 XX
 DB 1646 -----VSRGT-----NANNTVTVTQDGTSEVTVP-----VKHVI 1677
 QY 674 GGVGFTSGSRVYLASY--NGDOPRGNFOPFLYVGYGLGYOQTRGCFWYGYKLNN 731
 XX
 DB 1678 -----PEIVASHYTVQGDFFPAGN-----GSSASDFPKLSNG 1710
 QY 732 SPYDVLDSPRVGTETNOFRRTSLTPYVAGYLEEGAR--SFSNTPYIAGQDT----- 783
 XX
 DB 1711 S-DIADA-----TIIM-VSGQAPKNKDNFRIGEDIVTAHILIDGETTPITK 1754
 QY 784 -----PESRSTFQSGYSD--NTVEIOSV-----LGFD----- 809
 DB 1755 TATYKVVRTVPKHVEFETARGVLYPVSVDMDYKQVYKPNVNSMSTNACHMNFQVGTGYP 1814
 QY 810 -----GIRNNL-----NVGKASFLNSNRPNGLEMIATYLSQIGLARTSGLPN 858
 DB 1815 NKDVVGISTRLIRVTDNRQTDLTLSKVCKPDPRIANSVT-----YKAGLTN 1864
 QY 859 QQ-----PFGTHOYISVSPGDOSIKNIRTIPTG-----NQ 891
 DB 1865 QEIKVNNVNNSSVYLFKADNPPLANTNTHSGSFSSVYTVSDALPNGIKAKSSISMNN 1924
 QY 892 LMYLFTFNENKSSVYTLRLADSSNDPSSFSPTSLDVNEIGVILPLDINSFYTMAA 951
 DB 1925 VTY---TTQDEHGQVYVTRNESVDSNDATVTVTPOLATTEGAVFIKGDDG---DF 1977
 QY 952 GNVALFSSNP-----GSPGSY--TAVNTPNOLSDIAFEGSGAKTSDFWGTIOFK 1000
 DB 1978 GHVERFIQNPFGATVAMHDSPTMKNTVNTIKTAVVILP--NGQTR-----NVEVP 2029
 QY 1001 PDEVYIQTGFTSQVARNFVTVNOSFLNSLVDFTPANAGT--NRYVAVDDGNL----- 1050
 DB 2030 VKVYVYVAV--AKAPSRDVKGQNLTN-----GTDAMN--ITFDPNTNNGITAM 2075
 QY 1051 TNQNLPLKYO--IQYL--DCKYPAKILKNNLVTFESYNFGALPSWVVPYTAIGSTLGILA 1106
 DB 2076 ANROOPNNOAGVOHLNDVTVPGISAKKRPVAVNYQF--EFPQTYTTTGVGT----- 2129
 QY 1107 IMILGLAIGLPLRAORLQD-----KGEKTFTEKVDYLT--AAVSYVKKIITQYAN 1158
 DB 2130 -----LASQOASGYAHMONATGLPTDGFYKNNRPJTGTNDNMSAMKPNYAKVYNA 2183
 QY 1159 KKKPALGAGSGDKKPAAPAKAPAPKAPKASSPAKPTGP 1201
 DB 2184 KYDVIYNGHTFATSLPAKFVYKDVOPAKPYTETPAAGAITIAP 2226
 XX
 RESULT 15
 AAE00701
 ID AAE00701 standard; protein; 2123 AA.
 XX
 AC AAE00701;
 XX
 DT 02-JUL-2001 (first entry)
 XX
 DE Moraxella catarrhalis outer membrane protein-106 (OMP106).
 XX
 KM Outer membrane protein-106; OMP106; haemagglutination; vaccine;
 XX bacterial infection; immunogen; cytotoxic; antidiotic;
 XX passive immunisation.
 XX
 OS Moraxella catarrhalis.
 XX
 PN US6214981-B1.
 XX
 PD 10-APR-2001.
 XX
 PF 12-NOV-1997; 97US-0968685.
 XX
 PR 03-MAY-1996; 96US-0642712.
 XX

PA (ANTE-) ANTEN BIOLOGICS INC.
 PT Tucker K, Plosilla L, Tillman UF;
 XX
 DR WPI: 2001-281002/29.
 DR N-PSDB; AAD04029.
 XX
 PT Novel nucleotide sequences encoding Moraxella catarrhalis outer
 PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
 PT infections and as vaccine against Moraxella catarrhalis infection of
 PT mammals -
 XX
 PS Claim 7; Column 53-64; 49pp; English.
 XX
 CC The present sequence is haemagglutinating Moraxella catarrhalis outer
 CC membrane protein-106 (OMP106). The OMP106 is used as a therapeutic
 CC and prophylactic vaccine against M. catarrhalis infections of mammals.
 CC It is used for diagnosis of bacterial infections and as reagents for
 CC clinical or medical diagnosis of M. catarrhalis infections and for
 CC scientific research on the properties of pathogenicity, virulence and
 CC infectivity of M. catarrhalis. It is also used as a probe to identify
 CC the presence of M. catarrhalis in biological specimens and to identify
 CC other bacteria that encode a polypeptide related to M. catarrhalis
 CC OMP106. OMP106-derived polypeptides are used as ligands to detect
 CC antibodies elicited in response to M. catarrhalis infections and also
 CC as immunogens for inducing M. catarrhalis-specific antibodies which are
 CC useful in immunoassays to detect M. catarrhalis in biological specimens.
 CC Cytotoxic antibodies are useful in passive immunisations against
 CC M. catarrhalis.
 XX
 SQ Sequence 2123 AA:
 3.2%; Score 205; DB 22; Length 2123;
 Query Match Best Local Similarity 19.88; Pred No. 0.0011;
 Matches 286; Conservative 157; Mismatches 502; Indels 498; Gaps 72;
 QY 24 SASPGKQSDSKNDNTQVNOARTL-----DANSVALAGLQGN-----GSLFVTVARD 71
 DB 482 AATYG--QLKKVQNTASALQTFYKVKVKNQNDANDSILITVGKKNKPDGTQVNT----- 535
 QY 72 VDDNFTITRANGTIKIDLSFTKPLYGDLSDDCGKVKQIVSDYTSRNRFDQRTAVY 131
 DB 536 -----LKL-----KGENGVTVTETNG-----TYVE 556
 QY 132 ALLVNDENAVHLKRIINTNSRIGRNNNSKFVIG-----GYDNPAAHYR 175
 DB 557 GLNQNGGLTVGNSLTINDGLSVKNTNSNKOIQVGADGIFTTDSNKRPGAGIEN---TTR 613
 QY 176 FTDDGTRKNTFNOQGEIVNDFIDA--PILPKDLHPWYNYIQRKILPNDVNTAVY 232
 DB 614 ITRDG--IGFANNIGS-----LDANKPRLTPTGINAGKEL-----TWOSAIP 656
 QY 233 WPGVHVSQTNDADGMEFCGNGQIT--NDPIAQTITTD-----NONPSTFNSGAMPANR 287
 DB 657 -----ATNG--GOLDFMNRILSTANTEKSGSAATIDLYNLSOVLPTFGDGPVNTYK 707
 QY 288 Y-----DSQLVNKHRIKTSFQIDERINTNSNRIGNR- 318
 DB 708 LGEILIKVKGCTTADDLTKNNIGVAVDSTDSLVK-LAKTLDLD-AVNTKTLTJASDKV 765
 QY 319 -----NNNSKF-----VIGGVNPAHYIRFTD-----DGTFN 346
 DB 766 TVDSGNNTFAKLQNGDLTFGSKONTGATPATNSKTIYGVG---LKFDDNNGIALDGTTY- 820
 QY 347 FTNQTGEIVNDFILD--APILPKDLHPWYNYI--QRKILPNDVNT---AVWPVYGRV 400
 DB 821 ITRDKVGFAPKQSDSLDKSKPYLDK-----KLKVGVEITTINGINAGKAKITGASNTLT 874
 QY 401 SGTNADG-----MFCGNGQITND--PIAQTITTDNONPST 437
 DB 875 DATNATGHTQGIYDSTDKTRAASIGVLNAGFNLKNNGAKDFVSYDTIVD----- 928
 QY 438 FNSGAMPANNRDLSQLVNKHRIKTSFQDEKFTVYEWMTGSEENKRIITLRTAGSLPSNER 497

Db 929 FINGNATTAQVYDCKAS-----KVAYDVNDGTTIHLTGADGNKNGIGKTTTLTKT-- 981
QY 498 YMTLDIPGTPQYTLK-----EDSVNFSRLXNSVNSLSFGDSIYIFGISELPSLMYYSF 553
Db 982 ---DAKGKAINFESVNSGDDKALINAKDIADNLNTLA--GEIRNRTKGTAD----- 1026
QY 554 PTRLSDTLALNGYKTDIDIEASSPTDNGCTTNGTTTADTSSGSTAGTGNTNTSOTVSNP 613
Db 1027 -----TALQTFQYKVK-----ENGDDN-----DADTIT-----VGDAKTNQ----- 1060
QY 614 TLTNYSFGIDSKPTSAKIDETNMADPNVIEARIYAEYRLGIONEIPITNAGNFIKNTI 673
Db 1061 -VNTLKLKGNGLDIQTNK-----DGVTEGINTQSGLKAGNNTTLNNGLSIKNTA 1111
QY 674 G-----GVGFTSGRSYVLARASYNDQDPTGNFOPFLYFGLGIOQTRGT--FW 722
Db 1112 GNEQIQVGADGVFAKVNNGV--GAGIDGTRITRDEIGFAGTNGSLDKSKPHLSKDGIN 1170
QY 723 YGTYKLNNSPYVDLDSPRVGETNQFRRTSLTPVMGGL-----TE-----EGARFSF 772
Db 1171 AGGKKTIN-----IOSGELAQNSND-----AVTGKTIYDLKTELENKISSTAKTAQ 1216
QY 773 NTPYIRAGDTPESRSIFQSGYSNDNTYEQIYSLGF--DGIRNNLN-----VGKASSF 824
Db 1217 NSLHEFSVADDEGNNTFVSNPSSYDTSKTSVDYTFAGENGITTKVKNGVVRVGIQDTKG 1276
QY 825 LNSNR---PNPNGLEMIATTYIRSQIGLARTSGLPNOQPFCTTHOVISVSPGDDQFSIK 881
Db 1277 LTTPKLTVGNNGKGIY-----IDSONGONTITIGLSN-----TLA 1311
QY 882 NIRTIFPGNQLWTFLETFENENKSSVYTLR---LADSNPDASSFSPTSILIDVNEIGVI 937
Db 1312 NV-----TNDKGSVRTTEOGKIIRKDEKTRAA-----SIVDYLSAGFN 1349
QY 938 L---PLLD--NSFTYVNAAGVNALEFSSNFGSPSYTAVTFENQNLSDIAFEGSGAK--- 988
Db 1350 LOGNGEAVDVFVSTYDVFNEFADGNATAKYTDTSKTSKVYDVNVDDTIEVKDKKLG 1409
QY 989 ---YTSDFWGTIQFNPDEFLIONGFT-----SOVARNFVYNQSFNSLVDFTPANAG 1037
Db 1410 KTTTLTSTGTGA-----NKFALSNQATGDALVYKASDIVAHLNTLSGDIQTAKGASQANS 1464
QY 1038 TNYRVVDDPGNLTNQLPLKVOIQYLDGKYYDAK-----LKNNNLVTFESYNNFG 1087
Db 1465 AG---VDADGN-----KVIYDSTDNKIYQAKNDGYDKTEYAKDKLVAAQOTPDG 1513
QY 1088 ALPSWVVPYPAIGS-----TGILAIMTIL 1111
Db 1514 TLAQMNKSVYINKEQVNDANKKOGINEDNAFVKGLEKASDNKTKNAAYTVGDLMNV--- 1570
QY 1112 GLAIGLPLAQRKLO---DKGFTTFPKV--DTLTAVGSYKKIITQ-----T 1155
Db 1571 -----AQTPPLTFAGDTG--TTAKKLGELTLIKGGQTDINKLTDNNIGNVAGTDF 1619
QY 1156 ANYKRRPAALGAGKSG---DKKPAAPAPAPAPKSPAPKSPAPTGPK---SGAPT 1207
Db 1620 VKIAKDLTNLNSVNAAGTKRIDKGVSPVSSGQAKANTPYLSANGLDLGKVISNWGKGT 1679
QY 1208 KPT 1210
Db 1680 KDT 1682

Search completed: October 10, 2003, 16:33:25
Job time : 73 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 16:35:22 : Search time 271 Seconds
(without alignments)
727.753 Million cell updates/sec

Title: US-09-901-572A-4
Perfect score: 6413
Sequence: 1 MNISKRLKSYTLIGLAVFG.....APTKEPTAPKPAKPTAPKE 1224

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6413	100.0	1224	US-09-901-572A-4	Sequence 4, Appl1
2	6413	100.0	1224	US-10-131-591A-25	Sequence 25, Appl1
3	5201	81.1	1042	US-10-131-591A-78	Sequence 78, Appl1
4	223	3.5	2893	US-09-883-227-522	Sequence 522, App
5	205.5	3.2	2122	US-09-813-214A-9	Sequence 9, Appl1
6	205.5	3.2	2344	US-09-815-242-12713	Sequence 12713, A
7	197.5	3.1	5795	US-09-815-242-12610	Sequence 12610, A
8	197	3.1	1778	US-10-238-075-749	Sequence 749, App
9	196.5	3.1	2150	US-10-135-322-17	Sequence 17, Appl1
10	192.5	3.0	2283	US-10-173-502-4	Sequence 4, Appl1
11	192.5	3.0	2468	US-10-246-330-4	Sequence 4, Appl1
12	192	3.0	1536	US-10-193-764-63	Sequence 63, Appl1
13	190	3.0	2710	US-10-011-366-6	Sequence 6, Appl1
14	189.5	3.0	2353	US-09-797-862-33	Sequence 33, Appl1
15	189	2.9	1536	US-10-092-880-2	Sequence 2, Appl1

16	183	2.9	1600	US-10-092-880-10	Sequence 10, Appl1
17	180	2.8	1723	US-09-841-132-394	Sequence 394, App
18	179	2.8	1723	US-09-841-132-395	Sequence 395, App
19	178.5	2.8	1349	US-09-815-242-5898	Sequence 5898, App
20	178.5	2.8	1349	US-09-815-242-5898	Sequence 13137, A
21	177	2.8	1833	US-10-175-275-4	Sequence 4, Appl1
22	177	2.8	1833	US-10-175-282-4	Sequence 4, Appl1
23	177	2.8	1932	US-10-175-275-3	Sequence 3, Appl1
24	177	2.8	1992	US-10-175-282-3	Sequence 3, Appl1
25	174.5	2.7	1781	US-09-995-749A-2	Sequence 2, Appl1
26	174.5	2.7	2086	US-09-815-242-5639	Sequence 5639, App
27	173.5	2.7	1848	US-09-839-996-6	Sequence 6, Appl1
28	173.5	2.7	1848	US-10-080-505-6	Sequence 6, Appl1
29	172	2.7	1230	US-09-881-732A-150	Sequence 150, App
30	171.5	2.7	1073	US-10-193-764-45	Sequence 45, Appl1
31	171.5	2.7	1079	US-10-193-764-43	Sequence 43, Appl1
32	171.5	2.7	1599	US-10-092-880-9	Sequence 9, Appl1
33	171	2.7	6281	US-09-815-242-12996	Sequence 12996, A
34	167.5	2.6	1974	US-09-895-913A-12	Sequence 12, Appl1
35	167	2.6	1477	US-10-193-764-67	Sequence 67, Appl1
36	166.5	2.6	1095	US-10-193-764-65	Sequence 65, Appl1
37	166	2.6	3472	US-10-029-120-4	Sequence 4, Appl1
38	166	2.6	3472	US-10-027-806-4	Sequence 4, Appl1
39	166	2.6	3472	US-10-034-623-4	Sequence 4, Appl1
40	166	2.6	3472	US-10-027-801-4	Sequence 4, Appl1
41	165.5	2.6	820	US-10-156-761-7990	Sequence 7990, App
42	165	2.6	1075	US-09-801-368-110	Sequence 110, App
43	165	2.6	1270	US-10-245-802-16	Sequence 16, Appl1
44	165	2.6	5935	US-10-243-243A-8	Sequence 8, Appl1
45	164.5	2.6	1220	US-10-193-764-28	Sequence 28, Appl1

ALIGNMENTS

RESULT 1
US-09-901-572A-4
: Sequence 4, Application US/09901572A
: Publication No. US20030165534A1
GENERAL INFORMATION:
: APPLICANT: Nippon Zeon Co., Ltd.,
: TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
: FILE REFERENCE: J209
: CURRENT APPLICATION NUMBER: US/09/901, 572A
: CURRENT FILING DATE: 2003-03-11
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 1224
: TYPE: PRT
: ORGANISM: Mycoplasma gallisepticum
: FEATURE:
: OTHER INFORMATION: MGC3 encoded by mgc3 gene
US-09-901-572A-4

Query Match 100.0%; Score 6413; DB 12; Length 1224;
Best local Similarity 100.0%; Pred. No. 0;
Matches 1224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNISKRLKSYTLIGLAVFGALGSASFGRKSDKSDNTQVYQARTLDANSVRLAGLQ 60
|||||
Db 1 MNISKRLKSYTLIGLAVFGALGSASFGRKSDKSDNTQVYQARTLDANSVRLAGLQ 60
|||||
QY 61 NSLSFTVLRVDVDDNITTAANGTITIKLDSFTKPLYLGLSDDCGGYKVKQVSDYTTSRN 120
|||||
Db 61 NSLSFTVLRVDVDDNITTAANGTITIKLDSFTKPLYLGLSDDCGGYKVKQVSDYTTSRN 120
|||||
QY 121 REFQROTARYVALLVNDENVHLKRTINTSNRNGNNNSKFVIGVDPAHYIRPTDGG 180
|||||
Db 121 REFQROTARYVALLVNDENVHLKRTINTSNRNGNNNSKFVIGVDPAHYIRPTDGG 180
|||||
QY 181 TKFNFTNOTGELVNDFFILDAPILPKDLHPDWYNTLYIORKILPNDVNTAVPWPVGRVSG 240
|||||

Wed Oct 15 11:17:05 2003

us-09-901-572a-4.rapb

161 TKFNFTNOTOGEIYNDPILDPKDLHPDWNLXIQKILPNDVNTAVVMPVGRVSG 240
241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFENFTNOTOGEIYNDP 360
301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFENFTNOTOGEIYNDP 360
361 LDAPILPKDLHPDWNLXIQKILPNDVNTAVVMPVGRVSGTNADGMDGCGNGOITNT 420
361 LDAPILPKDLHPDWNLXIQKILPNDVNTAVVMPVGRVSGTNADGMDGCGNGOITNT 420
421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERFVYEMTSEE 480
421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERFVYEMTSEE 480
481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVSLSFIDSLIYF 540
481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVSLSFIDSLIYF 540
541 GTSLEPSLWYSEFTRLSDLTALNOYKTDIEASSTDNCTTNGTTTADTSSGSGAGT 600
541 GTSLEPSLWYSEFTRLSDLTALNOYKTDIEASSTDNCTTNGTTTADTSSGSGAGT 600
601 GNTTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONE 660
601 GNTTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONE 660
661 PITNAGNFIKNTIGVGTSTGRVYLRSYNGDORPTGNFOPFLYVFCYLYGOOTRGT 720
661 PITNAGNFIKNTIGVGTSTGRVYLRSYNGDORPTGNFOPFLYVFCYLYGOOTRGT 720
721 FWYGTYKLLNNSPDVLDSPRVGTETNOFRTSLIYPVMGYLTLEGARSSNTPIYRAQ 780
721 FWYGTYKLLNNSPDVLDSPRVGTETNOFRTSLIYPVMGYLTLEGARSSNTPIYRAQ 780
781 GTPESRSIFOGSYDNTYEXIYQVILGFDGIRNNLNVGVAASSFLNSNRPNGLEMI 840
781 GTPESRSIFOGSYDNTYEXIYQVILGFDGIRNNLNVGVAASSFLNSNRPNGLEMI 840
841 TTYLSQIGLARTSGLPNOOPFGTTHOYISVSPDQFSSIKNITIPGQNLWYFLTNE 900
841 TTYLSQIGLARTSGLPNOOPFGTTHOYISVSPDQFSSIKNITIPGQNLWYFLTNE 900
901 NKNSSYYTLRLADSSNPDASSSFPTSLIDVNEIGVILPLDNSFTYVNAAGNALFSSN 960
901 NKNSSYYTLRLADSSNPDASSSFPTSLIDVNEIGVILPLDNSFTYVNAAGNALFSSN 960
961 PGSPGTYAVNTFNOMLSDIAFEGSGAKYTSDFMGTIOFKPDEVLIONGFTSOVARNFT 1020
961 PGSPGTYAVNTFNOMLSDIAFEGSGAKYTSDFMGTIOFKPDEVLIONGFTSOVARNFT 1020
1021 NOSFLNSLVDFTPANAGTNRVYVDPDGLTNQNPPLKVOIYLDGKTYDAKIKNNNLVT 1080
1021 NOSFLNSLVDFTPANAGTNRVYVDPDGLTNQNPPLKVOIYLDGKTYDAKIKNNNLVT 1080
1081 FSYNNFALPSWVPPAIGSTLGIILMIIILGALIGPLRAQRKLODDKFKTTFKVDTL 1140
1081 FSYNNFALPSWVPPAIGSTLGIILMIIILGALIGPLRAQRKLODDKFKTTFKVDTL 1140
1141 TAAVGSYKRIITOTANVKKRPAALGAGSGDKKPAAPAKPSAPKASPAKPTG 1200
1141 TAAVGSYKRIITOTANVKKRPAALGAGSGDKKPAAPAKPSAPKASPAKPTG 1200
1201 PKSGAPKPTAPKPAAPKPTAPKE 1224
1201 PKSGAPKPTAPKPAAPKPTAPKE 1224

Sequence 25, Application US/10131591A
Publication No. US20030059799A1
GENERAL INFORMATION:
TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
FILE REFERENCE: J209
CURRENT APPLICATION NUMBER: US/10/131,591A
CURRENT FILING DATE: 2002-08-15
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 25
LENGTH: 1224
TYPE: PRT
ORGANISM: Mycoplasma gallisepticum
FEATURE:
OTHER INFORMATION: MGC3 encoded by mgc3 gene
US-10-131-591A-25
Query Match 100.0%; Score 6413; DB 15; Length 1224;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MNISKRLKSYTLIGLAVFALGASFGKSDKSDNTQLVNQAFTLDANSVRLAGLG 60
1 MNISKRLKSYTLIGLAVFALGASFGKSDKSDNTQLVNQAFTLDANSVRLAGLG 60
61 NGSLEFNTVLRDVEDNEFTTAANGTIIKIDSEFTKPLXGLDSDDCGKYVKQIVSDYTSRN 120
61 NGSLEFNTVLRDVEDNEFTTAANGTIIKIDSEFTKPLXGLDSDDCGKYVKQIVSDYTSRN 120
121 RFQROTRAVYALVNDENAVHLKRIINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGT 180
121 RFQROTRAVYALVNDENAVHLKRIINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGT 180
181 TKFNFTNOTOGEIYNDPILDPKDLHPDWNLXIQKILPNDVNTAVVMPVGRVSG 240
181 TKFNFTNOTOGEIYNDPILDPKDLHPDWNLXIQKILPNDVNTAVVMPVGRVSG 240
241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
241 TNADGMDGCGNGOITNTDPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKT 300
301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFENFTNOTOGEIYNDP 360
301 SFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHVIRFTDGTGFENFTNOTOGEIYNDP 360
361 LDAPILPKDLHPDWNLXIQKILPNDVNTAVVMPVGRVSGTNADGMDGCGNGOITNT 420
361 LDAPILPKDLHPDWNLXIQKILPNDVNTAVVMPVGRVSGTNADGMDGCGNGOITNT 420
421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERFVYEMTSEE 480
421 DPIAOTKTTTNDONSTFNSGAMPANNRYDSOLNVKHKIKTSFOLDERFVYEMTSEE 480
481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVSLSFIDSLIYF 540
481 NKNITRLATGSLPSNERWILDIPTGPVYTLKEDSVNFSRLYNSVSLSFIDSLIYF 540
541 GTSLEPSLWYSEFTRLSDLTALNOYKTDIEASSTDNCTTNGTTTADTSSGSGAGT 600
541 GTSLEPSLWYSEFTRLSDLTALNOYKTDIEASSTDNCTTNGTTTADTSSGSGAGT 600
601 GNTTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONE 660
601 GNTTNTSQVSNPTLNTYRSFGIDSKPTSAKIDETNMADPNVIEARLYAEYRLGIONE 660
661 PITNAGNFIKNTIGVGTSTGRVYLRSYNGDORPTGNFOPFLYVFCYLYGOOTRGT 720
661 PITNAGNFIKNTIGVGTSTGRVYLRSYNGDORPTGNFOPFLYVFCYLYGOOTRGT 720
721 FWYGTYKLLNNSPDVLDSPRVGTETNOFRTSLIYPVMGYLTLEGARSSNTPIYRAQ 780
721 FWYGTYKLLNNSPDVLDSPRVGTETNOFRTSLIYPVMGYLTLEGARSSNTPIYRAQ 780

QY 781 GTPESRSIFOSGYSDNTYEYIOSVLGFDGIRNNLVNGVAKASSFLNSNRPNNGLEMTAA 840
|||||
Db 781 GTPESRSIFOSGYSDNTYEYIOSVLGFDGIRNNLVNGVAKASSFLNSNRPNNGLEMTAA 840
QY 841 TTYLRSQLGLARTSGLPNOOPFGTTHQVIVSVPDQSSIKNIRITFPGNQWLYFLEPNE 900
|||||
Db 841 TTYLRSQLGLARTSGLPNOOPFGTTHQVIVSVPDQSSIKNIRITFPGNQWLYFLEPNE 900
QY 901 NNKSSVYTLRLADSSNPDASSSPSLIDVNEIGVILPLDINSFTYVNAAGVAFESSN 960
|||||
Db 901 NNKSSVYTLRLADSSNPDASSSPSLIDVNEIGVILPLDINSFTYVNAAGVAFESSN 960
QY 961 PGSPGTYAVNTEFNQNLSDIAFEGSGAKYTSDFWGTIOFEPDEYLIONGFTSOVARNFTV 1020
|||||
Db 961 PGSPGTYAVNTEFNQNLSDIAFEGSGAKYTSDFWGTIOFEPDEYLIONGFTSOVARNFTV 1020
QY 1021 NOSFLNSLVDFTPANAAGTVRRVVDGMLTNQNLPLKVOIQYLDGKYYDAKLKNNNLVT 1080
|||||
Db 1021 NOSFLNSLVDFTPANAAGTVRRVVDGMLTNQNLPLKVOIQYLDGKYYDAKLKNNNLVT 1080
QY 1081 FSYNNFGALPSWVPVPAIGSTGLIAMIILGLAIGPLRAORIKLDGKFTTFFKVDTL 1140
|||||
Db 1081 FSYNNFGALPSWVPVPAIGSTGLIAMIILGLAIGPLRAORIKLDGKFTTFFKVDTL 1140
QY 1141 TAAVGSVYKKIITQTANVKKKPAALGAGSGDKKPPAAAKPAPAPKPSAKSSPAKPTG 1200
|||||
Db 1141 TAAVGSVYKKIITQTANVKKKPAALGAGSGDKKPPAAAKPAPAPKPSAKSSPAKPTG 1200
QY 1201 PKSGAPTPTAPKPAAPKPTAPKE 1224
|||||
Db 1201 PKSGAPTPTAPKPAAPKPTAPKE 1224

RESULT 3
US-10-131-591a-78
; Sequence 78, Application US/10131591A
; Publication No. US2003005979A1
; GENERAL INFORMATION:
; APPLICANT: Nippon Zeon Co., Ltd.,
; TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
; FILE REFERENCE: J209
; CURRENT APPLICATION NUMBER: US/10/131,591A
; CURRENT FILING DATE: 2002-08-15
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 1042
; TYPE: PRT
; ORGANISM: Mycoplasma gallisepticum
; FEATURE:
; OTHER INFORMATION: Modified MGC3 antigen (M11-BTR)
US-10-131-591a-78

Query Match 81.1%; Score 5201; DB 15; Length 1042;
Best Local Similarity 84.7%; Pred. No. 0;
Matches 1018; Conservative 1; Mismatches 21; Indels 162; Gaps 1;

QY 23 GSASFGEKDSKNDNTQVLNQAARTLDANSVRLAGLQNGSLFNTVLRVDVDFTTAANG 82
|||||
Db 3 GSASFGEKDSKNDNTQVLNQAARTLDANSVRLAGLQNGSLFNTVLRVDVDFTTAANG 62
QY 83 TTIKLSFTKPLGLDLSDDCGYKVKQIYSDYTTSNRFDOROTRAYVALLVNDENAVH 142
|||||
Db 63 TTIKLSFTKPLGLDLSDDCGYKVKQIYSDYTTSNRFDOROTRAYVALLVNDENAVH 122
QY 143 LKRINTNSNRIGNRNNNSKEFVIGVDNPAHVIRFTDGTGFNFMTOTGEIVNDFIDAP 202
|||||
Db 123 LK----- 124
QY 203 ILPKDLHPDWYKLYIQARKILPNDVNTAVVWPVYGRVSGTNADGMEFCNGQITINDPIA 262
Db 125 ----- 124

QY 263 QTKTTDQNPSTFNSGAMPGANNRYSQNLVHKRIKTSFOLDERINTNSNRIGNRNNNS 322
125 -----RINTNSNRIGNRNNNS 140
QY 323 KFYIGVDNPAHVIRFTDGTGFNFMTOTGEIVNDFIDAPILPKDLHPDWYKLYIQARK 382
|||||
Db 141 KFYIGVDNPAHVIRFTDGTGFNFMTOTGEIVNDFIDAPILPKDLHPDWYKLYIQARK 200
QY 383 ILPNDVNTAVVWPVYGRVSGTNADGMEFCNGQITINDPIAQTCTTTDQNPSTFNSGA 442
|||||
Db 201 ILPNDVNTAVVWPVYGRVSGTNADGMEFCNGQITINDPIAQTCTTTDQNPSTFNSGA 260
QY 443 MGCANRYDSQLVHKRIKTSFOLDERFVYPEMTGSENNKQIRLATGSLPSNEHWYILD 502
|||||
Db 261 MGCANRYDSQLVHKRIKTSFOLDERFVYPEMTGSENNKQIRLATGSLPSNEHWYILD 320
QY 503 IPGTPQVTLKEDSVNFSRLYLSVNSLSFIDGSIYIFGTSELPJLWYSFPTRLSDTLA 562
|||||
Db 321 IPGTPQVTLKEDSVNFSRLYLSVNSLSFIDGSIYIFGTSELPJLWYSFPTRLSDTLA 380
QY 563 LNOVKTDIDTASSTDNCTNGTNTTADTSSGSTGAGTNTNTSQTIVSNPTLNTYRSFG 622
|||||
Db 381 LNOVKTDIDTASSTDNCTNGTNTTADTSSGSTGAGTNTNTSQTIVSNPTLNTYRSFG 440
QY 623 IDSKPTSAKIDETNMADPNVIEARIYAERYLGIQNEIPTTNAGNFIRNTIGVGFTSTG 682
|||||
Db 441 IDSKPTSAKIDETNMADPNVIEARIYAERYLGIQNEIPTTNAGNFIRNTIGVGFTSTG 500
QY 683 SRVYLASVNGDORPTGNFQPLVYFEGYLGQYQOTRTGTFWYGTYKLLNSPYDLDSPRV 742
|||||
Db 501 SRVYLASVNGDORPTGNFQPLVYFEGYLGQYQOTRTGTFWYGTYKLLNSPYDLDSPRV 560
QY 743 GTETNOFRRTSLIYPPWAGGLTEEGAGSEFNTPIYIRAQGTPEPSRSIFOSGYSDNTYEYI 802
|||||
Db 561 GTETNOFRRTSLIYPPWAGGLTEEGAGSEFNTPIYIRAQGTPEPSRSIFOSGYSDNTYEYI 620
QY 803 QSVLGFPGIRNNLVNGVAKASSFLNSNRPNNGLEMTAATYLRSQLGLARTSGLPNOQP 862
|||||
Db 621 QSVLGFPGIRNNLVNGVAKASSFLNSNRPNNGLEMTAATYLRSQLGLARTSGLPNOQP 680
QY 863 GTTHQVIVSVPDQSSIKNIRITFPGNQWLYFLEPNEKKSSVYTLRLADSSNPDASS 922
|||||
Db 681 GTTHQVIVSVPDQSSIKNIRITFPGNQWLYFLEPNEKKSSVYTLRLADSSNPDASS 740
QY 923 FSPSLIDVNEIGVILPLDINSFTYVNAAGVAFESSNPSGSGTYAVNTEFNQNLSDIAF 982
|||||
Db 741 FSPSLIDVNEIGVILPLDINSFTYVNAAGVAFESSNPSGSGTYAVNTEFNQNLSDIAF 800
QY 983 EGSAGKYSDFWGTIOFEPDEYLIONGFTSOVARNFTVNOFSLNSLVDFTPANAAGTVRRV 1042
|||||
Db 801 EGSAGKYSDFWGTIOFEPDEYLIONGFTSOVARNFTVNOFSLNSLVDFTPANAAGTVRRV 860
QY 1043 VVDPDGNLTQNLPLKVOIQYLDGKYYDAKLKNNNLVTFSYNNGALPSWVPVPAIGSTL 1102
|||||
Db 861 VVDPDGNLTQNLPLKVOIQYLDGKYYDAKLKNNNLVTFSYNNGALPSWVPVPAIGSTL 920
QY 1103 GILAIMIILGLAIGPLRAORIKLDGKFTTFFKVDTLTAAVGSVYKKIITQTANVKKP 1162
|||||
Db 921 GILAIMIILGLAIGPLRAORIKLDGKFTTFFKVDTLTAAVGSVYKKIITQTANVKKP 980
QY 1163 AALGAGSGDKKPPAAAKPAPAPKPSAKSSPAKPTAPKPSAGAPTPTAPKPAAPKPTAP 1222
|||||
Db 981 AALGAGSGDKKPPAAAKPAPAPKPSAKSSPAKPTAPKPSAGAPTPTAPKPAAPKPTAP 1040
QY 1223 KE 1224
||
Db 1041 KE 1042

RESULT 4
US-09-882-227-522
; Sequence 522, Application US/09882227

Publication No. US20030158396A1
 GENERAL INFORMATION:
 APPLICANT: Kleantous, Harold
 APPLICANT: Al-Garawi, Amal
 APPLICANT: Miller, Charles
 APPLICANT: Tomb, Jean-Francois
 APPLICANT: Oomen, Raymond P.
 TITLE OF INVENTION: Identification of Polynucleotides
 TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
 TITLE OF INVENTION: Encoding No. US20030158396A1 Helicobacter Polypeptides in the
 TITLE OF INVENTION: Genome
 FILE REFERENCE: 06132/047002
 CURRENT APPLICATION NUMBER: US/09/882,227
 CURRENT FILING DATE: 2001-06-15
 PRIOR APPLICATION NUMBER: US 08/902,615
 PRIOR FILING DATE: 1997-07-29
 NUMBER OF SEQ ID NOS: 638
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 522
 LENGTH: 2893
 TYPE: PRT
 ORGANISM: Helicobacter pylori
 US-09-882-227-522

Query Match 3.5%: Score 223; DB 12; Length 2893;
 Best Local Similarity 20.5%: Pred. No. 7.8e-07;
 Matches 281; Conservative 160; Mismatches 473; Indels 458; Gaps 73;

14 GGLAVPG-----ALGSASFGFQSDKSNCTQVNOARTLDANSYRLAGLQNSGL 64
 287 GGTYTFNKEFSATNNTFFSGSFNFKGVSSFN-GTFSNASTYFD-NQATFQNSFNGGT 344
 65 F-----NTVLRYDDNFTIANGCTIILKDSFTKPLGLDSDCCGKYKQVSD- 114
 345 FTENNQPTNNAQHPOIUNSSFGSNATLTK-----GVNFQQAFFNS 387
 115 ---YTSRNRDQOTRAYIYLLVNDANHLKRINSNRIGNRNNSEFYIGY---- 167
 388 NHQLTIONASFNNAFTNNTGKITTEKDSFN---NTFTSYDTNNMS--VAGVTLGG 441
 168 -----DNPAHVRFTDGTKEFNTN-----OTGELYNDPILDAPI--LP 205
 442 KNDLNGSTLDGSSKITLA-QGTFFNLISGSEKSYTLINSSGITYSLMLNHNGLT 500
 206 KDLHPWYNYLQKILNDVNTAVVPPVGRVSGTANADGCMFCGQITNTDPIAOTK 265
 501 SALKTN-----ESLSNPOSFAOGLMDITYNGV-----TGLQLN-ENATSK 541
 266 TTIDNQNPTFNSGAMPANRRYSQNLVVKHRI-KTSFOLDERINTNSNRI-----GNRN 319
 542 PIDSSPKSSTNSTQV-----YGVYKIGDTIKYKLOETSHNSIITIOALESSTYT 591
 320 ---NNSKRVYIG-----VDNPAHVR---FTDDGKREFTNOTOGEIYND--- 358
 592 PPVINGSKFDLASNYINADWPYDHTKTYIKSONTFESGTYIPLSQVINGSTNSPK 651
 359 -FLIDAPILPKDLHPM--YNYLQKILNDVNTAVV---PPVGRVSGTANADGCMF 410
 652 TFSANGSNLYIGNSWTVDHNVSSGTVSFGDTSGALNGHCPMPYQCGTGT--NCTY 709
 411 DC-----GNCQITN-----TDPIAQKTTTDNONS-----TENSAMP 444
 710 SAYHYITANLRSGNRIGTGGAANLIFNGVDSINIANATITQHNAGITSSMTFTQSMD 769
 445 GANN-----RYSQNLVVKHRIKTSFOLDERINTNSNRI-----GNRN 500
 770 NSQNLNGINSNGKISVGTFTFNAQDKRIFNAGQAVENTNFN----- 814
 501 LIDIGTPQVTLKEDSVAVFSRLYN-----SYNSLSIGDSIYIFGSELPLM 549
 815 ---GGSYQ--FSGSLNFSNNQNSGFEISAKNNSFNANNSASFENNSATITSF 869
 550 YSEPTRLSDITLALNOKTKDIDEASFDNGTTNGTTTADTSSGSGTGAGTNT----- 603

870 VGDEFNANSNL-----QI-----AGNAVFENSTNGSONTANFNNGSVNISGNATFDNVY 919
 604 ---TNTS-----QTVSNPT---LNTYRFG-----IDSKP-----TSA 630
 920 ENGPTNISVKQOVLNNTLTKLNAPLSFGDGTTFNANSHVINIASITNGNTILVSS 979
 631 NKIDETNADPNVTEARLYAEIRGLONEIPITNAGFINNTGCVGTSGSVYLRAS 690
 980 KEIEYNNAFSKNLMOLINYQGH--GASEKLVSAGNGVYDV-----XS 1022
 691 YGDDRPNGNPOFL-----YVEGYLG-----YQRTTG--TWYGYTK- 727
 1023 FNNQ---TYNDEVSQNSISIRLGVNMVDEYDMKSDHLYQNALGEMTPNPNYN 1079
 728 ---LANSFY-----DYLDSPRVOTETNORRRTSLYPVAGYLTGEGARSPNRYIR 778
 1080 NLGNANNTIYYDKSIDFYASGTLTFAESQT-----FTGNSAIVFGAKSIRT 1130
 779 AGDPRERSIFQSGYSDNT-----YE-----YIQ----- 803
 1131 SLSDAPQSWTILRFQ--DKKAGASNDASGHCWNLQICGFTIGHYEAOKIYTGSIESGR 1188
 804 ---SVLGFDIRN-----NLNVYKASS--FLNSRPNPNCLEMIATYLR 846
 1189 ISSGAGASLNFNGLGILLTNATLYNRAAGTSSMNFIS-----NANIOAONSFYID 1242
 847 QIGLARTSGLPNOQPGTTHQVIVSPGQFSSIKNIRIIFPNQNLWFLFTENKSS- 905
 1243 D--TAONGGNPN--FSFNALNIDFNSNFERGVGTQSVFKNKNAISTFTNSLSG 1297
 906 YVTLR-----LADSSNPDASSFSPTSLIDVNEIGVY-----LPILDN 943
 1298 LYOMQARSVLEDSNLSVSVG---TSSIKANALNLSQNASINASHSTLELOGDNLVNDT 1354
 944 SFY-----TYNAAGNVAL--FSSNPQSPGTYT---AVN-----TFNQLSDIAFEGS- 985
 1335 SSLMNGSTINVSNNATINDYASLISNGSHLNGAVNENSANITSLNNSYVFKGAV 1414
 986 --GAKY-----TDFNGTQIOFKPDEVLIQNGFTSQVARNFYVNOSEFLNSLVDFPPANAG 1037
 1415 SLGGQFNLSSNLSLDFQGS-----SAITSNAFHYVDN-AFSQSPTFHQA--- 1459
 1038 TNYRVYVDDGNLTNOMLPLKVOLOYLDGKRYDAKLNKNNLYFYSNNGAL 1089
 1460 LDKAPLPLGNNLNPN-----NSSVLD-----LKNSQLV--FEDQSL 1496

RESULT 5

US-09-813-214A-9
 Sequence 9, Application US/09813214A
 Patent No. US20020177200A1

GENERAL INFORMATION:

APPLICANT: Tucker, Kenneth
 APPLICANT: Plosila, Laura
 TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE.
 TITLE OF INVENTION: SEQUENCE AND USES THEREOF
 FILE REFERENCE: 7969-089-999
 CURRENT APPLICATION NUMBER: US/09/813,214A
 CURRENT FILING DATE: 2000-03-20
 PRIOR APPLICATION NUMBER: 08/968,685
 PRIOR FILING DATE: 1997-11-12
 NUMBER OF SEQ ID NOS: 20
 SOFTWARE: PatentIn version 3.0

SEQ ID NO 9
 LENGTH: 2122
 TYPE: PRT

ORGANISM: Moraxella catarrhalis
 US-09-813-214A-9

Query Match 3.2%: Score 205.5; DB 10; Length 2122;
 Best Local Similarity 19.5%: Pred. No. 9.6e-06;
 Matches 282; Conservative 156; Mismatches 502; Indels 507; Gaps 70;

QY	24	SASGEFQOSKSNNTQLVNQAFLL-----DANSRLAGLSON-----GSLNPTVLAD	71
Db	482	AATG--QLKVNOFAESALQTFYVKVDKNGDANDSKITTVKNNKPDGTQYNT----	535
QY	72	VDDNFITPANGTIIKLDSPKPLVGLDLSDDCGYKQOIVSDYTTSRNRPDQROTAVY	131
Db	536	-----LKL-----KGEINGVDTTEITNG-----IYTF	556
QY	132	ALLVDEANVHLKINTNSNRIKRRNNSKEVIG-----GYDNPAYIR	175
Db	557	GLNOMNOLVTGNSLTINDGLSVKTKNTNSKQIQVADGITEPTDISNSKRGAIEN---TTR	613
QY	176	FTDQGTKEFNFPNOQGEIYNDFILDAIPLRDLHPWYNLXIQSKLII.PNDVNTAVPWPV	235
Db	614	ITRDG--IGFANNTGS-----LDAN-----KPLTPTGIN-----A	642
QY	236	GRVSGTNADDMGPFDCNG-----QITWDEIAOTKTTD-----NONSTFNSGAMP	282
Db	643	GGRKELTVQSAIINPATNGQOLDPEMNRILSTATERSGSAATIKDLNLISQVLEITPAGDGP	702
QY	283	GANNRY-----DSOLNVKRIKTSFOLDERINTNSNR	314
Db	703	NWTKKLGEILVKKGKTTADDLTFRKNNIGVADSTDSNLSYK--LAKTISLDLD-AVNTKTLT	760
QY	315	IGNR-----NNSKFIYIG-----GYDNPAYIRFPD-----DG	342
Db	761	ASDKVAVYVDSGNNTAKLONGLDTSFKONTGATPATNSKITGVDG---LKFIDNNGIALDG	816
QY	343	TKFNFNTOGOEIVNDFILD--ABILPRDLHPWYNLYI--ORKLIPNDVNT---AVYWP	396
Db	817	TTY--ITMDKYGFAQDSDLSKSKRYLKD-----KLKGEVELTTNGIINNGKAITGLS	869
QY	397	VGRVSGTNADG-----MFCGNGQITNTD--PIAQTITTTDNO	433
Db	870	NLTLDATNATGHTYQLGIVDSTDKTRAASIGDVLNAGFNLKNNGDADFVSTEDVD--	927
QY	434	NPTFNSGAMGANNRRDLSQNLVHRKIKTSQDDEKFPYPMPTSESEKNTIRLATGSLP	493
Db	928	---FINGNATTAAVYTDGRKAS---KVAADVVDGTHLHTLADAGKNNIGKVTYLLT	978
QY	494	SNERWYLDIPGTFQVYTLK---EDSVNVFSRLINSVNSLSFGDSIYIGTSELPLW	549
Db	979	KT-----DAKGDKAINFSVNSGDDKALINKADIANLNTLA--GEINTNGTAD-----	1025
QY	550	YSEPTRLSDLTALNOVKTDDIEASSTDNGTTTNGTTTADTSAGSGAGTGNNTTSOT	609
Db	1026	-----TALQTFQVKYK---ENGDDN---DADTIT-----YGDAKTNO-	1059
QY	610	VSNTPLMTYRBSFGIDSKRPTSAKIDEITNMADPNVIEARIYAEYRIGIONELPTINAGFI	669
Db	1060	-----VMTLKKKGNKGLDIOQNK-----DGTVTGELINQSGIKGAKNNNTLNNNGLSI	1106
QY	670	RNTIG-----GVGFTSTGSRVYLRSAYNGDORPFGNPOFLYVGYGYGOOTRGT	720
Db	1107	KNTAGNMQIQVAGACVAFKAKYNNGV--GAGIDGITTTRITRDRDIGRAGTNGSLDKSPHLK	1165
QY	721	--FWYGYTKLLNNSPYDVLDSPRVGTETNOEFRTSLAYPVNGYL---TE-----EGA	768
Db	1166	DGINAGKKKITN-----IQSGEIAQNSND-----AVYGGKIYDUKTELEMKISSTA	1211
QY	769	RSEFNTPIYIRAGQTPRESLIFQCGSYDNITYEIQSYLVG---DQIRNNLN-----GVK	820
Db	1212	KTAQNSLHEFSVADQGNFTVSNPYSYDTSKTSDYTFTEGNGEITITTKVKNKGVYRVID	1271
QY	821	ASFSFLNSNR---PNPNGLEMIATUTYLEISQIGLARTSGLPNQOPFGTHQVIVSPGQOF	877
Db	1272	QTKELTTPKLVGNNGKGIY-----IDSQNGQVTTIGLSN-----	1307
QY	878	SSIKNIARTIFPGNOLMYFLFTNENNNKSSVYTLR---LADSSNDPASSFSPTSILIDNE	933
Db	1308	-TLANV-----TJDKSGSVPTTEQGIKIKDEDKTRAA-----SIVDYL	1344
QY	934	IGVIL-----PLLD---NSFTYVNAAGNVALFESSNPGSPQSYAVNTPFNQNLDAFEESGA	987

[illegible]

```

RESULT 6
US-09-815-242-12713
; Sequence 12713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-32
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12713

3.2%, Score 205.5; DB 9; Length 2344;
Best Local Similarity 19.2%, Pred. No. 1,1e-05;
Matches 226; Conservative 167; Mismatches 483; Indels 301; Gaps 50;

```



```

Db 1564 TSGNTATTSFNVTAKP-----LRDKYRWGTS-----STAANPYRIANISMDA--1605
QY 325 VIGGVNDNPAHV--IRTDGDKFNFNTNQGEIVNDFILDAPLPLPDLHPDKYNLYIQRK 382
Db 1606 TVSOAQOTAIINSLMETETVPNKNMATAANETSXTV-----SNVSRGTNNANVTY 1658
QY 383 ILPNVNTAVVPMPCVRV-----SGTNADDMFCOGN-----414
Db 1659 VTYQDGTSTVTPVPAKHVPELVASHYTVQGDPPAGGSSHAD-YFKLSNGSAIPDXT 1717
QY 415 -----GQITNTPD-----PIAOKT-----TTDNONPSTFNGAM- 443
Db 1718 ITWVSGAENKDMTRIGEDINVTANILIDGETPIKTKATYKVSSVPKHVEFTNGAVF 1777
QY 444 PCANRKYDQLVNKKHKIKTSFOLDEKVFVPEWTGS-EKKNTIRLATGSLP---SNERTW 499
Db 1778 PGVSDVYDAKQYVK-PVNDSPWQOMAGRMFOETNSGSPKDVVGISTRIIRIYTYDHQIQ 1836
QY 500 ILDI-----PGTPQ-----VTLKEDSVNFSRLYLSVNSLSPIGDSIYFCTSELSLM 549
Db 1837 IKILAKVPPDPRIIDNSVTYKAGLTN--QOKIKNNVSSS-----STLFRKADNTP--- 1887
QY 550 VYSFPTRLSDLTALNQVKTDDIEASSTDNGTTTNGTTTAD-----TSSGSGAGGNTT 604
Db 1888 -----LITTN-----TTGSGNTAVVTVDALPNCVIRKARSITNNNTY 1926
QY 605 NTSQVYSNPTLNTYRFGIDSKPTSAKIDETNMADPNVTEARIYAE-----YLG----- 655
Db 1927 YTTODEHRAIDVTRNESVDSNDSAVTYTTPOLA--TTEGAVFKGDDGDFGVHEVR 1983
QY 656 ION-----EIRTNAGNEIRNTIGVGFTSGSRVLYLASVNGDOR-----P 697
Db 1984 IONPRGATVAMHNDNT-----WKNTVNTIKHTA-----VTLPSQOGRNVEPVKYP 2034
QY 698 TGNFO-PPLVYFGYLGQOTRTGTETWYGYTKLINSPYVLDSPRVGTETN----- 747
Db 2035 VANAKAPSRDVG--QMLTNGT-----DAINYTFD-----PNTWNGITAMANR 2078
QY 748 -----OFRRSIATYP-----VMGCVLLEGARSFSTPYIRAGDPEER 787
Db 2079 QOBNNOAGVOHLNDVTPGTTAKRVVYVNAVIOFE-----FPOTSTTYTGGT---- 2129
QY 788 SJFOSGSDNTYEXIQSVLGF--DGIR--NNLNVGVKASSFELNSRPNPGLMELIATY 842
Db 2130 --LANGTOASGYAHMQNANGLPDQFTYKMNNAATCTNDANMAAMKPN--AAKYVNAK 2184
QY 843 YLSQIGLARTSGLP-----NOQ-----FGTHOYISYSPGDOFSSIKNTITPPGNO 891
Db 2185 YDVIYNGHFFATSLPAKFVVKDQPAKPTVTETAAGAITTPG-----ANQTVNT-HAGNV 2239
QY 892 LMV-----ELFTNENKKS-----SVYTLR-LADSSN--PDASSSESPSLI 929
Db 2240 TTYADKLVIKRNNGNVYTTTRRNNTSPWKKEASATVAGIAGTNNQITTYAGTFNADTI 2299
QY 930 DV-----NEIGVILPLDINSFYTVNAGNVALFSSNPGSPGYSYAVNTF 973
Db 2300 QVAVATGSETISDEQRSDDFYVAPQPHANATKIMONGHIDITPNP-----SGHLINP- 2354
QY 974 NQNLISIAFE--GSAKYT-----SDFWGTIOKRPDEYLIONGFTSOVAKRPVTN 1021
Db 2355 --QOAMDIAYTEKMGNAEHSKITINVVGONNOM-TIANKPD-VYTLDAHKGK--TFENAN 2409
QY 1022 OSFLNSLVDFTPANACTNY-----RVVVDPPGNLTNQMLPJKYOI 1061
Db 2410 TYKPNASAITTP-KACTGHSASSNPSTLAPATHYVNTTEIYADYGSNVTAAEINNAV 2468
QY 1062 QYLDGKYYDAKLKN-----NNT-----VTFSYNN 1085
Db 2469 ----ANKRTATIKNGTAMPNTNLAGSTTTTLPVTVYND 2502

```

RESULT 8
US-10-238-075-749

```

; Sequence 749, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are 1st
; TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of 1
; FILE REFERENCE: BLANDINE
; CURRENT APPLICATION NUMBER: US/10/238,075
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 749
; LENGTH: 1778
; TYPE: PRN
; ORGANISM: Escherichia coli
; US-10-238-075-749

Query Match      3.1%; Score 197; DB 12; Length 1778;
Best Local Similarity 19.7%; Pired. No. 3.1e-05;
Matches 259; Conservative 150; Mismatches 475; Indels 428; Gaps 69;

2 NISKRLKSYTL-----IGLAVFGALGSASFQKSDKSN-----36
294 SIASKVNSTALGSNSLSGGENALALGEGSAGGSNSLAFSGOSRANGNSVAIGVAAAA 353
Db 37 -DNTQLVNQAFLDA-NSVRLAGLGONGSLFNTLRVDVDDNFTJANGTTIKL--DSFTR 92
Db 354 TDNSVALIGASTFTDASNTFVSGNSATKRRIVNNAAGAIINSTDAINGSQLYITSDVAK 413
QY 93 PLYGDLSDDCGGYKKQKVSDYDTTSRNRFDDQOTRAYALLVNDANHLKRNNSNR 152
Db 414 RLGG-----GATVGSQDGV-----TAVSYAL-----R 435
QY 153 IGRNNNSKFVIGGVNDNPAHVIRFTDDGTFE-----NFTNQ-----TOGEIYN 195
Db 436 SGTYYNVGD-ALSGIDN--NTLQMNKTAGAFSANGHANGNTKITTNAKGVSAVSTSDVYN 492
QY 196 DFLDAPILPKDLHPMYNLIOKILPNDVNTAVVPMYGRVSGTNADDGMFDCNGOI 255
Db 493 GSOL-----YDLQ--ODALLMN--GTA-----FSAHGTETASKITNTYTAGNL 531
QY 256 T--NTDPI--AOKTCTTDNONPSTFN-----SGAMPGAN-----NRYDSOLNV 294
Db 532 TAGSTDAVNGSOLKTTNDVNTYTTNTIATNTYITNLTDVAVNGLSDSLMKAKAFAFSA 591
QY 295 KHKIKTSFOLDERINTNSRIGRNNNSKFVIGGVNDNPAHVIRFTDDGTFKFNFTQOGE 354
Db 592 AHGTEATSKI-----TNVT-AGNLTAGSTDAVNGSQ-----LKTNDNVNTTNTIATNT 640
QY 355 IVNDFILDAPILPKDLHPMYNLIOKILPNDVNTAVVPM--PVGRVS--GTNADDMG 409
Db 641 TINTNLTLDA-----VNLSGDSLMLNKTAAFAAHGTDATSKI 679
QY 410 FDCGNGOIT--NTDPI--AOKTCTTDNONPSTFN-----SGAMPGANRVDOLNVK 457
Db 680 TNYTAGNLTAGSTDAVNGSOLKTTNDVNTNTNTINATNTNTINLTDVAVNGLDDSL- 738
QY 458 HRIKTSFOLDEKVFYDEWGTGSEENKNTIRLATGSLPSNERKYWILIPGPQVTLKDSVN 517
Db 739 NKTAGAFSAAH-----GTDATSKITNTVKAADLAGS-----TDVAVNGSOLKTTNDVNS 786
QY 518 VFSRLYLSVNSLSFGDSIYITIGTSELSPLMYSPPTRLSDLTALNOVKTDDIEASST 577
Db 787 TNYTNTINTLTDVAVNGLSDSLMN--KTAGAFSAHGTADATSKITNTVKAAGLTAGSD 842
QY 578 --NGT--TTNGTTTTADTSSGAGTGTNTYTSQTV-----SNP 613
Db 843 AVNGSOLKTTNDVNT-----NTNTITNLTSVGLKDKDSSLMKAKAFAFSAHG 892
QY 614 TLNTYRSFGIDSKPTSAKIDETNMADPNVTEARIYAEYRLGIQNEIPTNAGNFTIRNTI 673

```

us-09-901-572a-4.rapp

Db 161 EESVLTNOF-----PHNFLLSSNNNTLANNSSLIHTNOS-----VNSLNNN- 202

QY 244 DQGMDCGGQGTNTNDPIAQTCTTNDQNPSTFNSGAMPAGNRRYSQLTAVKHKRIKTSFQ 303

Db 203 -----NNO-TNOPIINNNNNNNNNN-----NNNSNSN----- 230

QY 304 LDERINTSNRIGNRNNNSKYFVGVDPAHVIFRTDQKRFPTQTOGETVNDPILDA 363

Db 231 -----NSNNNNNNNNNNNTI-----TDSPTYSKSHYET- NIGSHQRKKSIOSLIANS 277

QY 364 PI-LPKLDHPDWYNYLQIKIL-----PNDVNTAVVPPVGRVSGTADGMDGNGOI 417

Db 278 AHSFSK-----LKNKPLSSSTFSTVTC-----GAV 304

QY 418 TNTDPIAQTCTTNDQNPSTFNSGAMPAGNRRYSQLTAVKHKRIKTSFOLDEKFEVPEWG 477

Db 305 NN-----NSNNNNNNNNNSTGSLGAP-MDRSFDENIT-----ITESTG 344

QY 478 SEENKNTIRLATGS-LPSNERYWIIDIPTPO-VLKEDSVNYSR-LYNSVNSLSFIG 534

Db 345 ---GNNSPRNSGCSNGSG-----GIPLSPRLSLNGSVSPRNHILNNLN----- 391

QY 535 DSIYFGTSELPSL-WYSEPTRLDPLAQNQVKTDDIEAST-DNGTGTNGTTTADT 591

Db 392 -----NSNLPISPRINHINVSr-----LNNNNNNINPNNNNSNNNSNNVSPRNN 442

QY 592 SSGSGTAGT-----GNTTNTSGTQVSPT-LWTYSPGIDSKPTSAKIDETINAD 640

Db 443 NNINIPRNSISPRNSNGSGTTSIPRNSNNNNIINNINNNTLTPRPSPLEVANTN 502

QY 641 -PNVIEARIYAERLIGTONEPIITNACNFIIRNTIGVGFTSGSRVYLASTNG----- 693

Db 503 SPRLATSLNS-TPLYSLSLTSNNNOSNN-----NTNPSINNNGRNGCIOTI 552

QY 694 DQPTGNEQPELYVEGLGYOQRTGTGFWIGTYLL-----NSPQVLDSPRVGTE 745

Db 553 SEELIGN-KPYVYNNGN-NNNNNTTNNSTSNNTTNNNNNNNNNNNNNNNNNNNNNNNN 610

QY 746 TNOFRITSL-----TYPMGGYLTGEGARFSNTPYIRAGDTPESRSIFSOSGSDTYTX 801

Db 611 GNHSKNSLODEFTSMNGDSDISGAQS-GGSLRRRKKDDNDEN-----DGNSNNT----- 661

QY 802 IQGVLGFDGIRNNLNVGKASSFLSNRPN----- 831

Db 662 -----NSNNNSNN 716

QY 832 -----PNGLEMIATTYLRQSIGLARTSGLPNOQPF----- 862

Db 717 NNNNNNNNNNNYHNGATMMKSHNH-QSIGMSSSPKKNPKRFSRNCSLMGKRRAMAI 774

QY 863 -----GTHOVATSV-----SPGDOFSSIKNR-----T 885

Db 775 LGLETVSGSSISILATLVLRYSSENSIADDFARVADRTMLRIEENRLTYTQTLISLLS 834

QY 886 IEPG-----NOLM-----YFLPFNEENKS 904

Db 835 VEPSTSEQFVPFSKLSMDNAEGLEGIMAPRVSNLDRYTWEIHSVKIRIELYVNPNNSS 894

QY 905 SVYTLRLADSN-----PDASSP-----SPTSLIDVNE-IGVILPEL 941

Db 895 DKRDVPAAMAADYYPILFSEPOSDNHFKRYNIYSDMMRRPSLNTTRDGEKVSVAASYI 954

QY 942 DMSFTYV--NAGAVALEFSSNPGSPGYTAVNTFNQNLSDIAFEGSGCAVYTSDFPGTIOF 999

Db 955 -KLANVPRNSKSNVLL-----TYQAVYTYGKVL-----TYED 988

QY 1000 KPDEYTLNGFTSOVARNFY-----TNQSFNSLV--DFTP-----ANA 1033

Db 989 RRHEVI--GFASS-CREFISRMVSAQIRLTEEDSLDLYVFLDSTPIGELIYRASNA 1044

QY 1037 GTYRVAVVDPDGNLTNOLPLKVQIOYIDGKYIDAKLNNNNLVTFSGNNFGALPSPVV-- 109

Db 1044 GN-----DGSSTN-----IMNGKLEDR---SDMYTNTMNGV-RNMML 1083
QY 1095 -----PRAISSTIGLAIMIILGLAT-----GIPRA-----ORKIO 1126
Db 1084 RPSRKTNRKHYFFYPALIGVCMILSALVSFMPAVNTKRIKLSATNEDLHKEIYNRKLA 1143
QY 1127 DKGKTFKKVD 1138
Db 1144 EKALAESQERLE 1155

RESULT 10
US-10-172-502-4
; Sequence 4, Application US/10172502
; Publication No. US2003018583A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES.
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2283
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-4

Query Match 3.0%; Score 192.5; DB 12; Length 2283;
Best Local Similarity 18.7%; Pred. No. 9.9e-05;
Matches 219; Conservative 167; Mismatches 494; Indels 293; Gaps 47;

QY 2 NISKKKLSYTL-----IGLAVGALGASFGFGQSDK-----SNDWTOLV 42
Db 70 SISKKMTGGLKTTAVIGGAFYVNLHDOQ-AFAASDAPLTSELNQTSEVGNQNSTTIE 128
QY 43 NQARTIDANSYRLAGLQNGSLFNTRYLRVDYDNFTIANGTTIKLDSFTKPLYGLDLSDD 102
Db 129 ASTSTADSTSVTNSSSVOTSDTYSSEKSEKVTSTNSTSQOEKLTLS-----178
QY 103 CGGYKQAIYSDYTSRNRFDQRTARAYVALLVNDDEANVLKRIINTNSNRGNNSK 162
Db 179 -----TSESTSSKNTTSSSDTKSVASTSEOP-----INSTNSTSTNNSTQS 223
QY 163 VIGGVNPAHVIRETDGTFKFNFTQGEIVNDPILDKLHPDMVNLRYOKIL 222
Db 224 T-----TPSSVNLNKTSTSTLS-----TAVKLRT-----FSRLAMSTFA 258
QY 223 PNDVNTAVVPMVPGVGRVSGTNADGMDCCGQITNDPIAQTOTTTDNONPSTF--NSGA 280
Db 259 SAATTTAVT-----ANTTIVNKDNLKQYMTTSGN--ATYDOSTGI 296
QY 281 MGCANRRYDSQ-----LNVKRIKTSFOLDERINTNSRIGNRNNNSKFIYGVGDNPAPHVI 336
Db 297 VITLQOAYISQKGIITGTRIDSKNSKSFHSGKVLNKGKYECHGN-----GG-----341
QY 337 RPTDGTGFNFPTNQTQGE-----IVNDPILDAFILPKDLHPDMVNLRYOKIL 384
Db 342 -----DGIGFAFSPGVGTEGLNGAAYIGLSNAF-----GFKLDIYH-----NTSK 384
QY 385 PNDVNTAVVPMVPGVGRVSGTNADGMDCCGN--GOITNTPD--IAQTKTTDN--QNPSTFN 439
Db 385 PNSAKAKA-----NADPSNVAAGGAFGAFVYTDGVAATYTSSTADNMAKLN 432
QY 440 SCAMPGANRRYDSQLNVKRIKTSFOLDEKFFVPEWTGSEENKI-----484
Db 433 --VQPTNNTFQDDPIN-----INGDKTVMVTAKGOTWTRNISDWIAKSGTTNLS 482
QY 485 -----TRLA-----TGLSPSNERYWILDI--PGTPQVTLKEDSVNVFSTRYL--N 535

Db 483 MFASTGATNLDQVQGFTEFEYESAVTQYR--VDVTGKDIIPKPTYSGNVDQVVTIDN 540
QY 526 SVNSLSFTG-----DSIV-----IFGTSLEPLSMYVSFPPLRLSDIALNVO---566
Db 541 QQSALTKAGYNTSVDSYASTYNDTNKIVTKMNAQOSVYVYETPDVAKAPVTVGNQITIEV 600
QY 567 -KT-DDIEASSTNGTNTGTTTADTSSGSGTAGGNTNTSQTSNPLNTYRSPGID 624
Db 601 GKTMPNPIVLTDTDNGTIVNTVYTGILPSGLSDSDSATNIIIGTPTKIGOSTVYVSTDOAN 660
QY 625 SKPTSAKIDETMMADPNVI-----EARIYAERLQIONEIPIT-----NAGNFIRN 671
Db 661 NKSTTFITINVDTTAPVTPPIDQSEVVS-----PLSPIKIAIQDMSGNAVYN 710
QY 672 TI-----GGVGFSTGSRVVLKASVNDQRPCTGNFQPLVFGILGQQRRTGTFWGTIK 727
Db 711 TVYGLPSGLTFDSTNNTI-----SGTPTNIGTSTISLVSDASGNKTYTTF---KYE 759
QY 728 LUNNSPYDVLDSPRVGTETNQFRTSLT-----YPMGQVLTREGARFSMTPIRAQ 780
Db 760 VTRNS---MSDSVSTSGSTQOSQSVSTSKRADSQASTSTSGTIVSTASTSKSTSVSL 816
QY 781 GDTPEBSRIFQSGYSDNTYEQYISVLGFDGIRNNLVGVKASSFLNSRPNGLF---836
Db 817 DSVASAKSLSTESNSVSSSTSTSLVNSQSVSSMSDASAKSTSLSDSISNSSSTES 876
QY 837 -----MIATTVLSQIGLARTSGLPNOQPEGTHOYVSFGQFSSIKIRIPIFG 889
Db 877 LSTSTDSLRTSTSLSDSLMSRSGLSKQSLSTISGSSYSTASLSDSTSN--AISTS 934
QY 890 NOLMYFLFTNE--NKSSTVTLRLADSNDPSSSPSTSLIIVNEIGVILPLDMSFY 947
Db 935 TSLSEASSTDSISISINSASASTSKSDSOSTISLSTDSKSKMSTSESLDSTST 994
QY 948 VNAAGNVALFSSNPGSPGSAVTAVNTFNQNLSDIAFEGSGAKYSDFWGTIQFPRDEYLI 1007
Db 995 GSVSGSLIAQSQSVSTSTSDSMST--SEIVSD--SISGSLASDS--KSMVSSSMSTSQ 1051
QY 1008 NGFTSQ-VANFPTNOSFLNSLYDFTPANAGTN 1039
Db 1052 SGSTSESLSDSOSTSDSKLSQSTSGSST 1084

RESULT 11
US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; CURRENT FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2468
; TYPE: PRN
; ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4

Query Match 3.0%; Score 192.5; DB 12; Length 2468;
Best Local Similarity 20.2%; Pred. No. 0.00011;
Matches 308; Conservative 154; Mismatches 503; Indels 563; Gaps 84;

QY 61 NSGLEFNTVLRVDNFTTANGTIIKLDST---KPLYGLDLS-----DDCGGY 106
Db 559 DGTIVNVVARDAGNSSPPASVTVDAVAPATPTVDSNGTTLGTAEPGSSVTLTGNGN 618

1583 VISGTAEGATVTLTGTGNGNPIGOVYADSGGWSFTPGTPIANGSVINALAQDAAGNNS 1642
1583 VISGTAEGATVTLTGTGNGNPIGOVYADSGGWSFTPGTPIANGSVINALAQDAAGNNS 1642
925 PTLSDIVNEIGVLLPLLDNSFYVNA-----AGNVALFFSNPSPGSGTYAVNTFQNLSDI 980
925 PTLSDIVNEIGVLLPLLDNSFYVNA-----AGNVALFFSNPSPGSGTYAVNTFQNLSDI 980
1643 PTS-ATVDSLAPAPVATPDSNGSVIAGTAEGATVTLTGTGNGN-----IGOV 1689
1643 PTS-ATVDSLAPAPVATPDSNGSVIAGTAEGATVTLTGTGNGN-----IGOV 1689
961 AFESGGAKYTSDFWGTIQFKPDEXLIONGFTSQVAFANFTNOSF-LNSLVD-----F 1031
961 AFESGGAKYTSDFWGTIQFKPDEXLIONGFTSQVAFANFTNOSF-LNSLVD-----F 1031
1690 TADSGN-----W-----SFTGPFPLSNGTVVNAVADDAAGNISGVPSTVDVAVATPVI 1740
1690 TADSGN-----W-----SFTGPFPLSNGTVVNAVADDAAGNISGVPSTVDVAVATPVI 1740
1032 TPAN-----AGT-----NYRVV--DPDGNLTNOLPLKVOIOLGKYIDAKLKNNNLVTESY 1083
1032 TPAN-----AGT-----NYRVV--DPDGNLTNOLPLKVOIOLGKYIDAKLKNNNLVTESY 1083
1741 DPSNGVELSGTAEPGVRILYLDGNGNPIGQTL-----ADGS-----KTTTKKY 1137
1741 DPSNGVELSGTAEPGVRILYLDGNGNPIGQTL-----ADGS-----KTTTKKY 1137
1084 NNEGALPSWVPTAIGSTLIGLAIWILIGLAIPTLAQRKLDPKG-----QDPAGNTSGPASTVVDYV 1815
1084 NNEGALPSWVPTAIGSTLIGLAIWILIGLAIPTLAQRKLDPKG-----QDPAGNTSGPASTVVDYV 1815
1777 GMSWTFPG--PFLANGTVVNAVA-----ADGS-----KTTTKKY 1184
1777 GMSWTFPG--PFLANGTVVNAVA-----ADGS-----KTTTKKY 1184
1138 DPLTAAGSVYKKITTOFANVKKRPAAL-GAGK-----SGDKRAAAKPAAP----- 1175
1138 DPLTAAGSVYKKITTOFANVKKRPAAL-GAGK-----SGDKRAAAKPAAP----- 1175
1816 APATPIVINSNGSVIITGEVAGAKVILYLDGNGNPIGTEGTADSGNMTFTPETPIANGTVI 1875
1816 APATPIVINSNGSVIITGEVAGAKVILYLDGNGNPIGTEGTADSGNMTFTPETPIANGTVI 1875
1185 ---AKPSAPKASSPAKPTGPKGAPTKP 1209
1185 ---AKPSAPKASSPAKPTGPKGAPTKP 1209
1876 NAVAEDAGNAGSAPASTT-VDSVAPSPAP 1902
1876 NAVAEDAGNAGSAPASTT-VDSVAPSPAP 1902

Db

RESULT 12
US-10-193-764-63
Sequence 63, Application US/10193764
Publication No. US20030133943A1
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M.
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H.
TITLE OF INVENTION: PROTECTIVE RECOMBINANT HAEMOPHILUS INFLUENZAE HIGH
TITLE OF INVENTION: MOLECULAR WEIGHT PROTEINS
FILE REFERENCE: 1038-1239MIS
CURRENT APPLICATION NUMBER: US/10/193, 764
CURRENT FILING DATE: 2002-07-12
PRIOR APPLICATION NUMBER: 09/167, 568
PRIOR FILING DATE: 1998-10-07
NUMBER OF SEQ ID NOS: 91
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 63
LENGTH: 1536
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-193-764-63

Query Match 3.0%; Score 192; DB 12; Length 1536;
Best Local Similarity 19.5%; Pred. No. 5.9e-05;
Matches 266; Conservative 155; Mismatches 487; Indels 458; Gaps 66;

Db

27 FGKQSD-----KSDNTLOLVNARLTDAHSV-RLAG-LGQNGSLFNLYLRVDNDPITA 79
27 FGKQSD-----KSDNTLOLVNARLTDAHSV-RLAG-LGQNGSLFNLYLRVDNDPITA 79
105 FNIDQNEWVOLFQENNSAVFNR---VTSNQISQLGIDLSNGQVFL-----INFGITI 156
105 FNIDQNEWVOLFQENNSAVFNR---VTSNQISQLGIDLSNGQVFL-----INFGITI 156
80 ANGTTIKIDSFETKPLGLIDSDCCGKYKQVQIYSDYTSBN-REFDQRTA-----YIA 132
80 ANGTTIKIDSFETKPLGLIDSDCCGKYKQVQIYSDYTSBN-REFDQRTA-----YIA 132
157 GKDALLINNGFATSP--LDISNE-----NIK-----ANFEEQTKKALAEIVHGL 202
157 GKDALLINNGFATSP--LDISNE-----NIK-----ANFEEQTKKALAEIVHGL 202
133 LLYNDEANYHLKRITNSNRIGNRNNS-----KFVIGVDNPA-----HVIR 175
133 LLYNDEANYHLKRITNSNRIGNRNNS-----KFVIGVDNPA-----HVIR 175
203 ITVGKGSVNLIGKVKKNGVATSVNGGSSISLAGOKITISDIINPTITYSIAPENEAVN 262
203 ITVGKGSVNLIGKVKKNGVATSVNGGSSISLAGOKITISDIINPTITYSIAPENEAVN 262
176 FTVDGGRFNFOTOGELVNDIILDAPIPLPDLDHMDWNLVLTQRIKILPNDVNTAVWPV 235
176 FTVDGGRFNFOTOGELVNDIILDAPIPLPDLDHMDWNLVLTQRIKILPNDVNTAVWPV 235
263 LGDIFKGGINVRATIRNOCKLSADSVSKDSG---NIYLSAKKEGAETIGVI----- 314
263 LGDIFKGGINVRATIRNOCKLSADSVSKDSG---NIYLSAKKEGAETIGVI----- 314
236 GAVSGINADDMGMDCCNGOITNTNDPFAOTKTTTD---NONDSTENSGAMPANRRYSOL 292
236 GAVSGINADDMGMDCCNGOITNTNDPFAOTKTTTD---NONDSTENSGAMPANRRYSOL 292

Db


```

315 -----SAONQAGKGLMITGDKVTLKTGAVIDLSGKEGERTY-----LGDGERGCKK 363
QY 293 NKHKRIKTSFOLDERINTNSNRIGNNNSKFPVIGVDPAHAIKRTDGTCKNFNQ 352
Db 364 GIOLAKKTSLEKSTINVSKEKGR---AIWMDI-----ALIDG--NINAQGS 408
QY 353 GEIVND-----FILDAPILPKD---LHPDWNYIQRKILPNDVNTAVVWPWP 398
Db 409 GDIKTGFEVTSCHLFIKDNALVDKEMLLDPD-----NVSINMETAG 453
QY 399 RVSGTADDMEDCCGQIT---NDPIAQTKTTDNONSTFNSGAMPC--ANNR--YD 451
Db 454 R-SNTSEDEYTGSGSASTPKRKEKTYLTNTLE-----SLKSTFVNITANORIYV 508
QY 452 SOLANVHRIKTSFOLDEKFPVPEWT-----GSEENKNITRLATG-----SLPSMERW 499
Db 509 SNTNLSNGSLTL-----WSEGRSGGVEINNDIT---TGDDTRCANLTIYSGW 554
QY 500 I-----LDIPGTPVTLKED-----SVNV-----FSRLXNSVNS- 529
Db 555 VDVHKNISLGAOGNINIFAKODIAFEKGSNOVITGCGTISGNOGFRFNNSLNGTSG 614
QY 530 LSTIGSIYIFG-TSELPSLWYSEPTRLSDLTALNOVKTDIEASTDNGTNT---G 564
Db 615 LQFTTKRTNKYAITNKEGTLNISKVNISMVLPKKNESGYDKFKGRTYMLTSLNYESG 674
QY 585 TTTTADTSSGTCAGT-----GNTNTSQVYSNNTLNTYRSEFGIDSKPTSKANKIDETN 637
Db 675 ERFLLTIDSRGSDAGLTOPYNINIGISFNKDTTFENRNARVDFDKA-PIGINKYSSLN 733
QY 638 WADPNYIEARVAYEVLGIONEIPITNAGNFIRNTIGVGFTSGSRVLRASVNDQRP 697
Db 734 YASFN-----GNISVSGGSVDFT-----LLASSNVQRP 763
QY 698 TGNFOFLYVGYLGIOQTRTGTFWTGYTKLNNSEPYDVLDSRVTGETNQFRT----- 752
Db 764 G-----VVISKXYFNSTGSSLRFKTSKSTGTGFSIE 795
QY 753 -SLTFPVMGVYLTEGARSPSNTPYIRAOGDPRESISQSGSDMTYEVIOGVLPDG- 810
Db 796 KDLTLATGNTI---LLOVEGTIDGIGKGIYAKKNITTEGG--NITFGSKRAVTELEG 850
QY 811 --IRNNLNVGVKASFLNSNR-----NPNGL-----EMIAA 840
Db 851 VTINNANVTLLIGSDPDNHQKPLTIKKDVIINSGNLTAGNINYNAGNLVESNANPKAI 910
QY 841 TTYLRQIGLARTSGLPNOOPFGTTHQVIVSPGD-QFSSIKN-----IRTI 886
Db 911 TMTFVVGGLFDMKGNNSN-----ISIAKGARFKDIDSKNLSITNTSSSTYRTI 960
QY 887 FPGNQMLYFLFTNEN-----NKSYYTLRL----- 911
Db 961 ISGN-----ITKNKNDLNTNMGSDTEMOIGDVSQKEGNLTISSDKINITROITKAG 1014
QY 912 ADSNPDASSFSPTSLIDVNEIGVILLPLDNSF---TYVNAAGVALLFPSSNPGSPGY 967
Db 1015 VQENSDSDATNANLTIKTELKLTODLINSGFNAEITAKGSDLTIGNTN-SADGTN 1073
QY 968 TAVNTFNQNLSDIAFEGSGAKYTSDFWGTIOFKRPDEYLIQNGFTSOVARNFVTNOSFLNS 1027
Db 1074 AKKVTENO-VKDSKISADGHKVT-----LHS 1098
QY 1028 LVQFTPANAGTNRVAVDDPGNLNQNLPKVOIQVLDGKYYDAK--LKNNNLVTPSYNN 1085
Db 1099 KVTSSGNNTF-----DSSDNAGLTI-----DAKNVTYNNNTTSHKAVS 1139
QY 1086 FGLPSPWVP---TAIGSTLGILAIMIILGLAIGIPLRQQRKLODGKFTFKKVPDTLTA 1142
Db 1140 ISATSEITTKTGTTINATGNEIATAGNSIIG-----GIESSSQSV-TLTA 1186
QY 1143 -----AVGSYKKIITOTAN-----YKKRAALGAGKSGD 1172
Db 1187 TEGALAVSNISGNTVVTYANSALTTLAGSTIGTESVTTSSQSGD 1232

```

```

RESULT 13
US-10-011-366-6
; Sequence 6, Application US/10011366
; Publication No. US20030054493A1
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICIL-
; DISEASE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/011,366
; FILING DATE: 16-NOV-2003
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/957,310
; FILING DATE: 23-OCT-1997
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 24-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPBD-01121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-011-366-6

Query Match 3.08; Score 190; DB 15; Length 2710;
Best Local Similarity 19.38; Pred. No. 0.0002;
Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;

QY 42 VNQARTLDANSVRLAGISGLFNVYLRVDNDFITRANGTIIKLDSFTKPL----- 94
Db 1071 INNSLSIAATVAVISVIGALVITFLPIAGISGIBSLVNNELIHDKATSVNVPNHL 1130
QY 95 ---YGLDSDDCGGKRVQIVSDYTSNRNFDQRTAVYVALLVNDANVHLKRI 149
Db 1131 ESKRYGPKLREDD---KILVPIDDLVISEIDF-----NNN 1162
QY 150 SNRIKGNR-----NNSKFPYIGVDNPAHAIKRTDGTCKNFNTNQGELVNDPILDAPI 203
Db 1163 SIKLGTCNIIAMEGSGCHVTGNIDH-----FESSPSISSHIPSLSIYSA-- 1207
QY 204 LPKDLHPDWNYIQRKI--LPPDWNTAVVPMFVGRV-----SGTNADDMGFCGNG 253

```

Page 12

[illegible]

Db 1746 ---DPANSAGO-EVKAGDKVTEFKADNLIK---KSGKDPFYSLKKEIKLDLNS----- 1791
QY 569 DDEASSTNGT-----TTNGTPTTADTSSGSTAGTGNNTNTSO----- 608
Db 1792 ---VEFKDANGTGSESTKITKDGILTPPANGAGACAGANTANTISVTKDGISACNKAVTAV 1849
QY 609 -----TVSNPTL-----NTYRSF-----GIDSKPTSAKIDET----- 636
Db 1850 VSGLKFKGDGHTLANGTVADEFKHYDNAKDLNLNDEKADNNPTAADMTAAVGLDRGL 1909
QY 637 NW-----ADPNTEARITAEYRLGICQ--NEIPTNAGNFIKRTITGVGFTSTGSRVY 686
Db 1910 GWYISADKTTGEBN-----OENYNAOVNANNEVKFSKSG-----NGINVSAGKTLNGTTRY 1957
QY 687 LRASVNGDQRPNGNPFPLVPGYLGYYQ--TRTGFWYGTLYKLNSPYDVLDSRVGT 744
Db 1958 TRELAGEYVKSNEPT-----VKNADSETNLVKVGMYT-----SKEDIDPATSKEPMVG 2007
QY 745 ETNQPR-----RTSLTYPVMG-GYLTF-EEGARSFNTPYIRAQDTPESRSIF 790
Db 2008 KTEKYKVENGVKVSANGSKTEVTLTNKKSGYVTCNGVADALAKSGFELGLADAELAKAF 2067
QY 791 QSGYSN--TYEYIOSVLCFGDGR--NNLNVGKASSFLNSRPNPENGLEMTAATYLS 846
Db 2068 AESAKDKOLSKDAEAEVNAHDKVRFANGINTKVSATVEST--DANGDKV--TTTFVKT 2122
QY 847 QIGLARTSGLPNQPGFT--HQVTSVSGDQFSSIKNRTITFPNQMLYFLETENNK 903
Db 2123 DV-----ELPLTQITNTANGNKIVKKADGK-----WYELNAGGTAS 2159
QY 904 SSVYTLRLADSSNPDASSFSPSLIDVIEGIVLPLDNSFTYVNAAGVAFSSNPGS 963
Db 2160 NKEVTL-----GVADVANGK--KVKVTENGA-----DKWYTT-NADG----- 2193
QY 964 PGSYAVNTFPNQLSDIAIEGSGAKYTSDFWGTIOKPPDE-----YLIONGFTSOVARN 1017
Db 2194 ---AADKTKG-----EVSNDKYSTDEKHVYRLDPNNQSGKGVVIDNVANGELS-- 2239
QY 1018 FVTNQSFLNSVDFTPANAGTNYRVVVDGNTLNQLPLKVOIYLDCKYVYAKIKNN 1077
Db 2240 -ATSTAIINGSQLYAVAKGYTN-----LAGOVNNLEGGVNVKVG--KRAD 2280
QY 1078 LVTFSYNNFGALPSWVVPPTAIGSTLGLIAMIIL---GLAIGIPLRAORKLDK 1129
Db 2281 AGTASALASQLPQATMP---GKSMVAIAGSSVYGQNGLAIGV-----SKISDNG 2327
RESULT 15
US-10-092-880-2
; Sequence 2, Application US/10092880
; Publication No. US20020164354A1
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS OF NON-TYPEABLE
; FILE REFERENCE: HAEMOPHILUS
; CURRENT APPLICATION NUMBER: US/10/092,880
; PRIORITY FILING DATE: 1997-04-01
; PRIORITY FILING DATE: 1998-09-30
; PRIORITY FILING DATE: 1996-04-01
; PRIORITY FILING DATE: 1997-04-01
; PRIORITY FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1536
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-092-880-2

Query Match 2.9%; Score 189; DB 14; Length 1536;

Best Local Similarity 19.5%; Pred. No. 9, 8e-05;
Matches 266; Conservative 157; Mismatches 488; Indels 452; Gaps 67;
QY 27 FEGKSD-----KSNNTQLVNAOKTLDAHSV-RLAG-LGQNSLFNTYLRVDYDNPTRA 79
Db 105 FNIDQENWOFLOENNNSAVFNR--VTSNQSILGLIDSNQVPL-----INPGIT 156
QY 80 ANGTIKLDSFPPPLVGLDSDDCGGYKKQOIVSDYTSRN--RFDOROTRA-----YYA 132
Db 157 GKDALINTNGFTAST--LDISNE---NIK-----ARNFEDTKDAELAEIYNHGL 202
QY 133 LAYDEANVHLKRIINTNSNRIGNRNNNS-----KFVIGVDNPA-----HVIR 175
Db 203 ITYVKDGSVNLIGKVKNEGVLSVNGSISLAGQITISLDINPTTYSIAAPENEAVN 262
QY 176 FTDDGTKEFNFTNOTGEIYNDFLDAPILPKDLHPWYNLYIORKILPNDVNTAVPWPV 235
Db 263 LGDIFAKGNINVRATIRNOGKLSDSVSKDSG--NIVLSAKGEAEIGCVI----- 314
QY 236 GRVSGTNADDGMPDCNGOITNDPIAOKTKTTD--NONPSTFENGAMPANRRYDSOL 292
Db 315 -----SAONOAKGKMLMTTGDKVTLKTGAVIDLSGKEGETY-----LGDEREGKN 363
QY 293 NVKHKIKTSFOLDERINTNSNRIGNRNNNSKFVIGVDNPAHIREFTDGTKEFNFTNQ 352
Db 364 GIOLAKKTLEKSTINTVSGKEKGR--AIWGMGI-----ALIDG--NINAQS 408
QY 353 GEIYND-----FLIDAPILPKDLHPWYNLYIORKILPNDVNTAVPWPVGRVS 401
Db 409 GDIAGTKGFEVETSGHDLFIKDNAIVDA--KEV-----LLDFD--NVSINETAGR-S 455
QY 402 GTPNADGMPDCGSGQIT--NTDPIAOKTKTTDNONPSTFENGAMP--ANNR--VD SOL 454
Db 456 NTSEDEYVSGSGASATPKRNKKTTLTNTLE--SLIKKTEVNTANQRIYVNSI 511
QY 455 NVKHKIKTSFOLDEREVEYDEMT-----GSEENKNITRLATG-----SLPSNRYMI-- 500
Db 512 NLNSGSLTL-----WSEGRSGGGVEINNDIT--TGDDTRGANLTIYSGGVADV 557
QY 501 ---LDITGTPDYVLTAKED-----SVNV-----FSRLYLVNS-LSF 532
Db 558 HKNISLGAOGNINITKODIAFEKGSNOVITGOGTITSGNOKFRNNNSLNTGSGLOP 617
QY 533 IGDYIYFG-TSELPISWYSPFTRLSDLTALNOVTDDEASTNGTGTN---GTT 587
Db 618 TTKRTNKYALTNKFEGTLNIGKVNISWVLPKNEGSDKFKGTYNNLISLVNSEGEPN 677
QY 588 TADTSSGSTGAGT-----GNTTNTSQTIVSNPTLNTYRSFGIDSKPTSAKIDETNMAD 640
Db 678 LTIDSRGSDSAGTLQPYNLNGISFPMKDDTFENRANAVFDIKA-PGINKYSLNLYAS 736
QY 641 PNVEIARITAEYRLGIONEIPITNAGNFIRNTIGYGFSTGSRVYLRASVNDQRP 700
Db 737 FN-----GNISVSGGSVDFT-----LLASSSVVQPPG-- 764
QY 701 FQFELVVFYLGYYQVTRGTGFWYGTLYKLNSPYDVLDSRVGTETNOFPRT-----SL 754
Db 765 -----VVISKFFNVSTGSLFKTSIGKTKTFSTIEKDL 798
QY 755 TYPVMGYLTREGARFSNTPYIRAQDTPESRSIFQSGYSDNTYEYIOSVLCFGD--I 811
Db 799 TLNATGAGNT--LLQVEGTDMIGKIVAKNITEGEG--NITPFSRAVTEIEGVTI 853
QY 812 RNMLANGVKASSFLNSRNP-----NPNGL-----EMIAATY 843
Db 854 NNNANVTLLIGSDNDNQKPLTIKKDVIINSGLTLAAGINVIAGNLTVESNANPKATINP 913
QY 844 LRSGIGLARTSGLPNPOPFGTTHQVIVSVPGD-QFSISKN-----IRTIFFP 889
Db 914 TFNVGGLFDKNGSN-----ISIAKGARFADIDNSKMLSTITNSSSYRKLIS 963
QY 890 NQMLYFLFTNEN-----NKSYYTLRL-----ADS 914

us-09-901-572a-4.rapb

Wed Oct 15 11:17:05 2003

Db 964 N-----ITKNGDLNITNEGSDPTENQIGDVSOKEGNLTISDKINTKQITIKAGVDG 1017
QY 915 SNPDASSFSPTSLIDVNEIGVILPLIDNSF---YTVNAGNVALFSSNPGSGSYTAV 970
Db 1018 ENSDSDATNNANLTIKTELKLTQDLNISGFNKAETARKDSDLTIGNIN--SADGTNKK 1076
QY 971 NTFNOLSDIAFEGSGANYTSDFWGTIOFKPDEYLIONGFPTSOVARNFVYNOSFTLSLYD 1030
Db 1077 VTFNQ-VKDSKISADGHKV-----LHskVE 1101
QY 1031 FTNANAGTNRVVDEPDLNLTNQLPLKVOIQYLDKYYDAK--LKNNLVTFSYNNFGA 1088
Db 1102 TSGSNNTF-----DSSDNNAGLTI-----DAKNVTYNNNITSHKAVSISA 1142
QY 1089 LPSWVP---TAIGSTLGLIAMILIGLIGIPLRNORLQDKGFKTFKVDLTA--- 1142
Db 1143 TSGETTTGTGTINATGNVEITNAGTGSILG-----GISSSGSV-TLTPATEG 1189
QY 1143 --AVGSYYKKIITOTAN-----YKKPAPALGAGKSGD 1172
Db 1190 ALAVSNISGNTVTYANSALITTLASSTIKGTESVTTSSQSGD 1232

Search completed: October 10, 2003, 16:45:49
Job time : 281 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 16:32:21 ; Search time 23 Seconds
(without alignments)
2251.673 Million cell updates/sec

Title: US-09-901-572a-4

Perfect score: 6413
Sequence: 1 MNISKKLKSYTLIGLAVG.....APTPTAPKPAKPTAPKE 1224

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/prodata/1/laa/5A_COMB.pep:*
3: /cgn2_6/prodata/1/laa/5B_COMB.pep:*
4: /cgn2_6/prodata/1/laa/5A_COMB.pep:*
5: /cgn2_6/prodata/1/laa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/laa/backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	328	5.1	1637	1	US-07-665-792E-9
2	252	3.9	10182	4	US-09-134-001C-3159
3	224.5	3.5	2504	4	US-09-328-352-5821
4	205	3.2	2123	3	US-08-968-685A-10
5	199	3.1	1612	1	US-08-169-927-2
6	192	3.0	1536	4	US-09-206-942-67
7	190	3.0	2710	1	US-08-480-604A-6
8	190	3.0	2710	2	US-08-405-496A-6
9	190	3.0	2710	3	US-08-915-136-6
10	190	3.0	2710	4	US-08-957-310-6
11	190	3.0	2710	4	US-10-011-366-6
12	189.5	3.0	2353	3	US-09-377-155-33
13	189.5	3.0	2353	3	US-08-913-942-4
14	189.5	3.0	2353	4	US-09-669-974-33
15	189.5	3.0	2736	4	US-09-252-991A-30227
16	189	2.9	1536	1	US-08-038-682-2
17	189	2.9	1536	1	US-08-302-832-2
18	189	2.9	1536	2	US-08-530-198-2
19	189	2.9	1536	2	US-08-469-880-2
20	189	2.9	1536	2	US-08-728-470-2
21	189	2.9	1536	2	US-08-617-697-2
22	189	2.9	1536	3	US-08-719-641-2
23	188	2.9	2411	4	US-09-268-347-36
24	187	2.9	1529	3	US-08-728-470-10
25	187	2.9	1529	3	US-08-719-641-10
26	185	2.9	2354	4	US-09-268-347-47
27	183	2.9	1600	2	US-08-617-697-10

28	182.5	2.8	2314	4	US-09-268-347-49	Sequence 49, App1
29	180.5	2.8	2048	4	US-09-268-347-48	Sequence 48, App1
30	178.5	2.8	804	4	US-09-328-352-5545	Sequence 5545, App
31	177	2.8	1833	4	US-08-621-944A-4	Sequence 4, App11
32	177	2.8	1833	4	US-08-945-567D-4	Sequence 4, App11
33	177	2.8	1992	4	US-08-621-944A-3	Sequence 3, App11
34	177	2.8	1992	4	US-08-945-567D-3	Sequence 3, App11
35	175	2.7	2057	4	US-09-499-203-2	Sequence 2, App11
36	174.5	2.7	1278	4	US-09-604-957-3	Sequence 3, App11
37	173.5	2.7	1848	3	US-08-296-791-6	Sequence 6, App11
38	173.5	2.7	1848	5	PCT-US95-10661A-6	Sequence 6, App11
39	172.5	2.7	1287	1	US-08-200-232-2	Sequence 2, App11
40	172.5	2.7	1287	5	PCT-US95-02219-2	Sequence 2, App11
41	172.5	2.7	1287	5	PCT-US95-02219A-2	Sequence 2, App11
42	172.5	2.7	1335	4	US-09-134-001C-3716	Sequence 3716, App
43	172	2.7	1231	4	US-09-071-035-420	Sequence 420, App
44	172	2.7	1265	4	US-09-071-035-418	Sequence 418, App
45	171.5	2.7	1073	4	US-09-206-942-49	Sequence 49, App1

ALIGNMENTS

RESULT 1
US-07-665-792E-9
; Sequence 9, Application US/07665792E
; Patent No. 5281694
; GENERAL INFORMATION:
; APPLICANT: Baseman, Joel B.
; APPLICANT: Su, C. J.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: PRODUCTION OF MYOPLASMAL ADHESINS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: ARNOLD, WHITE & DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5-1/4" DSDD Diskette
; COMPUTER: Compaq (IBM PC Compatible)
; OPERATING SYSTEM: MS DOS 3.31
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/665, 792E
; FILING DATE: 19910607
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/118,967
; FILING DATE: NO. 5281694. 10, 1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/588,886
; FILING DATE: July 27, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/616,111
; FILING DATE: NO. 5281694. 21, 1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/697,349
; FILING DATE: May 8, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Melinda Patterson
; REGISTRATION NUMBER: 33,062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 787-1592
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1627 Amino Acids
; TYPE: AMINO ACID
; TOPOLOGY: Linear

5.18; Score 328; DB 1; Length 1627;

Db 1198 ATFNALPEHPNALAEQVSVVEASAYKPNPTSSGQGTOSTNSPYLLHLYKPKKVTQSDKLD 1257

QY 857 -----PNO-----OPGTTHOVISVSPGDFSSIKNIRITFPGNOLMYLFTNENKKS 904

Db 1258 LKNILDPNOVKRYKLRQSFQTDH---STQPOQ- -SLKTTTPEVFG-----TSSGNLS 1303

QY 905 SVYTLRLADSSNDPASSFSPTSLIDVNEI-----GVILLPLDINSFYTVNAGNVALEFSN 960

Db 1304 SVLS-----GGAGGSGSGGGGVDLSPEYKVGSLMGQLPST-----SDGNIS--STN 1351

QY 961 PGSGGSTYANTVFNQ-----NLSDIAFEESGAKYTSDFWGTIOFKPDEYLLIONGFTSOYA 1015

Db 1352 NLAENT-----NTGNDVYGVGRLS-----ESNAAKMNDVGIVRTPLAELLDGEGTADTG 1403

QY 1016 RNFTY-----NOSFLNSLYD-FTPAN-----AGTVARY-----VYPD 1046

Db 1404 PQSVKFKSPQDIDENRRLFTHPVTDLPDPVTMLVYDQITPLEFDIPASVPKMYRLKVLJSF 1463

QY 1047 DGNLTNOLPLKAYOIOYLDGKYYPDAKKNN-----NLVFS-----YNFCA 1088

Db 1464 DTN-----ESGLGLLEFFKP-----QDTQNNNNVQVNPNGDFPLPLTASOGPQTLFSPNO 1518

QY 1089 LPSWVVPFPAISGTGLIAMIILGLAIGIPRAQRKLODGFKTFEKKYDTLTAAGSVY 1148

Db 1519 WDGYVLPALATPIVIVIVLSTGLAIGIPMKKKOKALKGAFALNSQKVDVLTAKVGVF 1578

QY 1149 KKLITQGTANVKKKPPALGAGKSGDKKPPAAAKPAPAKPSAPKASAPKPTGPKSAPRK 1208

Db 1579 KELLNRT-GISQAPRL-----KQTSAAKPGAPRPVPKPGAKPPVQP-----PKK 1625

QY 1209 P 1209

Db 1626 P 1626

RESULT 2

US-09-134-001C-3159

US-09-134-001C, Application US/09134001C

Sequence 3159, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134, 001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064, 964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055, 779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO: 3159

LENGTH: 10182

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3159

Query Match 3.9%; Score 252; DB 4; Length 10182;

Best local Similarity 20.4%; Pred. No. 8.2e-08;

Matches 284; Conservative 148; Mismatches 407; Indels 556; Gaps 73;

Db 1398 QAAAKNSNR-----SNINSNEKOLAERGSGNGYSKSLIRDDGKSYLNSN----- 1443

QY 91 TKPLYLGLDSDDCGGKKVQIOISDVTTSRNRFDQRFRAYALLVND EANVHLKRITNS 150

Db 1444 --PIQVLDLVEPDNGGGGRQV-----SHS 1465

QY 151 NRIGNRNNNSKFEYIGGVNDPAHVIRFT--DDGTFKFNFTNO--TQGEIVNDFIDAPILPKDL 208

Db 1466 NVIYNENKNS- IYNGVQPEANGASAFENIDKVVYKAAANNGTIMGVITYKQVLIAPSPG- 1523

QY	209	HPDWNLYLXIO-----KILPND-VNTAVPMPVGR-----VSG-----T	241
Db	1524	-----YIEKLGONLSNTNNVINVTVFVPSDKVPST-----VGNVDHFFHYSGEETKNTI	1573
QY	242	NADGMEPCGNGOQTTNDPIAQTKTITDN-----ONPSTFNS-----GAMPANNRYD	289
Db	1574	NVND-----NYGLNTPVASTSDSAIYMTTRNNNELVGCAPVYJNSINKIVYKATDKSGNESIV	1630
QY	290	S-----OLANKHRIKTSFOLDERINTNSRIGNRNNNSKFVIGVDNPAHVIFRTDD-G	342
Db	1631	SFTVNIKPIKNEKYRITSSS-----NOTPVARSIMQNNANLSIEDQNRVKSSLSMTILG	1685
QY	343	TKFEVNOTOGELIYNDPILDPILPKDLHPWYLYIQRKILPNDVNTAVPMPVGRYSG	402
Db	1686	TR-NYVESN-----NDVSIOYVS-RVNR-SG	1709
QY	403	TNADGMEPCNGOITNT-----DPIAQTKTTDNONPSTFNSGAMPANNRDS	452
Db	1710	NNAIYVNTTTFSDGTTNTIIVPKHVILLEVPTRTYRGRQ-----FPRGKTSNDPFS-	1765
QY	453	OLANKHRIKTSFOLDERKVFYPEWTGSEBN-----KNITRLA	488
Db	1766	-----LRTGPGVDARIWVWVNOGPDINSNOIGRDLTHAEIFEDGETTPIRKDTYKL	1818
QY	489	TGSLPSN-----EGYW-ILDI-----PGTPO	508
Db	1819	SOSIPKQIETTINGRENSGSDAPGNFVQAVNDOYIPEHMFPRMAQSGGTPSSRNAGSET	1878
QY	509	--VLT-----KEDSVNPFSLYLS--VNSLSFI-----GDSIYIFGTSLSPLMYSSF	553
Db	1879	KTVYVVOGNGTEVANNVLFKYPKPKPVIDSNSVSKOLNQOILVRVNPVNAOYVTLXO-	1937
QY	554	PTRLSDLTALNQVKTDDIEASSTDNGTTTNGTTTADTSGST-----GAGTGNTNTSQ	608
Db	1938	-----SNGFIYIPMTNTTIDNSGIAVYTIQGLEPTGNITAKTS	1974
QY	609	TVSNPPLNTYSFEGIDSKPT-----SAKKIDET-----NMADPN	642
Db	1975	MTNNVYTYTKONSSGLASNTTEDISVFENSEDOVNVYTAGMOAKNDGIKINGTNNYFNDFN	2034
QY	643	VIEARIYAEYRLIGIONE-----IPIT-----NAGNF	668
Db	2035	SFISNIPIRASHLTWNNEEPNSMKNNIGTTKIVYTLPRHQSTRYVDIPTIITYPVYTAKNP	2094
QY	669	IR-----NTIGVGFTSTGSRVLYRASNGDORPTGNF--QP	703
Db	2095	VRODKGRNLNGTDVYNYIIEFNNNRLGCT-----ASMKDORQRDKNKIAGQN	2142
QY	704	FLVYFYGIG-----YQOTRIGTF--WGYIKLLNS--PY	734
Db	2143	LIALVNPFGISTPLEVPYKVVWVNFDEPTQPIYKIOVGCTPFPGTWAGYKHLHENGEGAPI	2202
QY	735	D-----VLDSPRVGTETNO-----FRRTSL--TYPPW-----GGYLTBEGAR--SF	771
Db	2203	DGMKFTYMQOSTGTSDQOSLAIYTRIPFVYTGTYDVVNPBNNMGWQTSOSAKFTVTNAK	2262
QY	772	SNTPIYRAO--GD--TPES-RSIFOSGSDNTEYIOSVGFDFGIRNNLWVYKASFL	825
Db	2263	PNOPTITGSKTGDTVTPPGAVARNLIS--TNDYIDA--SADKIYINKN-GNKLITTFV	2315
QY	826	NSN-----RPNPNCLEMIATATYLRISOIGLAFRTSGLPNOQPFGTHOIVISVPGQD	876
Db	2316	KNNQGRMWTVEGSPDINGI--GPTNNGTAISLS-----LAVRPGD-	2354
QY	877	FSSTKNTFTTIPGQNLWAFLETTNENKNSVYTLRLADSN-----PDAS-S	921
Db	2355	--STEALATBEGGETI-----SFSVSEIYIVAKPOEQUATHTYDNGTFDILPDNSRN	2406
QY	922	SFSPSTLIDNVEIGIYILPLDNSFYTVAAAGNAV--LESSNPGSPGASYAVN-----TFN	974
Db	2407	SLNPTEREIN-----YTEKLNENEFQKSTFIIRKNNNGKATTINKKNRYVEFN	2453
QY	975	ONLSDIABEGSGAKYTSDFMGTIOQKPEDEYLIQNGEFTSQVARNFVYTNOSFLNSLVDTPPA	1034

```

Db      2454  QDNCKVVFSAHTIKPNSOI--TITPKAGQGCTEN-----TNEPTIQA-----PA 2495
Qy      1035  NAGTNYRVVYDDPDG-NITNOMLPLKVOIQYLDGKYYPDAKLNNMLVTFSYNNFGALPSWV 10933
Db      2496  QHTLTINEIVEEGQANTNDNDINNAQV-----PNKRRAVIAKGSN--ALPT-- 2539
Qy      1094  VPTAIGSTGLGILAIM 1108
Db      2540  -NLAGGSTSHIPVVI 2553

RESULT 3
US-09-328-352-5821
; Sequence 5821, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINET
; TITLE OF INVENTION: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5821
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5821

Query Match      3.5%; Score 224.5; DB 4; Length 2504;
Best Local Similarity 18.8%; Pred. No. 8.8e-07;
Matches 267; Conservative 179; Mismatches 442; Indels 529; Gaps 68;

Qy      34  KSNNDTOVNOARTLIDANSVRLAG-----LGG-----NGSLFNT-VLRDYDNI- 77
Db      876  KTELTNTKIDNTKTELEKGLNFPAGNSGADVHRKLGDKLNIYGAASAIPPAKTSGETNITT 935
Qy      78  -TAANGTII-----KIDSFTR-----PLYGLDLS-----DCGGYKVOI--VSD 114
Db      936  RTQDQIGIELLKDSKSPDSVTGNTLTNTLGTITKEGSIKQGINNG--KQITNVAD 992
Qy      115  YTSRNFEDQOTRAYIYALVYDEANVHLKRIKNTS-----NRIGKRN--NNSK 161
Db      993  GINAKDAVNVHQ-----LTKRVENLNGRTITDNNQNDAKKDLGNADJNTNNLNAK 1045
Qy      162  FVIGGVDPNAVIRFDTGCKFNT-----NOTGEIYNDF-----ILDAP1--LPK 206
Db      1046  KDLDG-----QITDTNTLNMTKQDLTTOINDTKTELNNITGNTKTELSKIDSTKT 1097
Qy      207  DLHPDMWNLV-----IORKI--LPNDVNTAVVPMVGRVSGTNA-----D 244
Db      1098  ELEKKGILNFPAGNSGADVHRKLGKELKNIIGGAASIPPAKTSGETNITRTQDQIGIELLK 1157
Qy      245  DGMFDCNGQITNTDPIAQTKTTTDNONSPFNSGAMGAN-----NRYDSQLNVK 295
Db      1158  DSKFD-----SVTTGNTLTNTLNGLTITKEGSIKQGINAGSKOITNVADGINAKDA--VNV 1212
Qy      296  HRIKTSFOLDIRIMTNSRIG-----NRR--NNSKFYIGGVDPNAVHYIRFTDD 341
Db      1213  QLTVEKENLNRITDNTNQLNDAKKDLGNQIADTNTKNNLDKKDIDG-----IQTDT 1264
Qy      342  GTFKNFT-----NOTOGEIYNDF-----ILDAP1--LPKDLHPDMWNLV-----IQ 380
Db      1265  NTKLNTYTDOLTQIINDTKTELNNITGNTKTELSKIDNTVTELENKGLNFPAGNSGADVH 1324
Qy      381  RKI--LPNDVNTAVVPMVGRVSGTN-----ADGGMFD----- 411
Db      1325  RKLGDKLNIYGAASAIPPAKTSGETNITRTTQDQIGIIELLKDSKSPDSVTGNTLTNTNG 1384
Qy      412  -----CGNGQITNT-----DPIAQTKTTTDO----- 433
Db      1385  LTIKEGSIKQGINAGSKOITNVADGINAKDAVANKSOLDNLAKQNAQTDAAVKAYDDAK 1444

```

434 -----NPS---T 437
1445 TDKYTLKDKDGTVDNKKAGHISSTSKAVNGSOLKHSINSIGNTVNPDSLT 1504
438 FNSGAMPANRRYSOLNVK---RIKTSFOLDEFFYPEMTGSEENK-----I 484
1505 TNNIGGTGNININDAISEVKNATAKTIVTEGDIVKVEYNKDGSGINYEATKKDLTL 1564
485 TRLATGSLPSENERWILLIPGTPOVYLKEDSVNPSRLYLSVNSLSTIGSIYFGISE 544
1565 NSVTTGDTYVNNNG--LTIKDGPSTL--KDGINASGKOITTVADGIN--AKAVNDOLTK 1619
545 LPSLWYYSFPTRLSDL--TALNOVKD-----DIEASDNGTGTNGTT 586
1620 VKE-----NLNGRITDNNOLNDARKDLGNOLADTNKNLNDAKDKLGNOTDNTNTLNTK 1675
587 TTDATSSGSGAGTNTSOTVSNPTLNTYRSFGIDSKPTSAKIDET-----N 637
1676 DQLTQINDTKTELNNTIGTKTELN-----SKIDSTKTELENKGIN 1717
638 WADPNVIEARLYAEYRLGIONEIPITNAGFIR-----NTIGVGFT-----STGSRV 685
1718 FAG-----NSGNDVHRKLGKELINTIGGAASATPAKTSGEV 1754
686 VLRSYNGDORPTNPQPFYVFGYLGYOQTRGTGTYGTYKLLNNSPYDVLDSRVGTE 745
1755 ITRTYODGIG-----IELKDSKEDSVTTGTNTLN 1784
746 TNOFRRTSLTYPVNGYLTTEG-----ARFSNTPYIRAGDTPES 786
1785 TN-----GLTIK-EGPSVTEKGINAGGKITTVADGVNAKADYKKSQDLNLAKONATD 1838
787 RSIFPGSYDNTYETIOSVLGFDG--IRNNLVGVKASSFLNSNPNGLEMAATYLR 845
1839 AAV---KDDAKTKRYLKLKGDGTVLDN---VKAGHISSTSKAVNGSOLHSINSIK 1891
846 SOIGLARTSGLEPNQOPG--TTHOYVISP---GDPSIKNITPTIPGNOLWFLEPTE 900
1892 NSIG-----GNFVNPDSLTNNIGTGKNNINDAISEVKNAT----- 1931
901 NKSSSY---TLRLDSSNPDASSFPSFTSLD--VNEIGVILPLDLSFTVNAAGNVA 955
1932 KAKTYTEGDNIYKETVKKDSTNYEYSTKDLNLNSVTGDTVLNNNGLTIDGDSI- 1990
956 LFSSNPGSPGTYAVTFNONISDLAFESGAKYISDPWGTIOFPEYLLONGFTSQVA 1015
1991 -----TKDGVNAGGKITDVA-----NGVIAQNS 2014
1016 RNFVNTOSFLNSLVDTPANAGTNYRVVDPDGNLTNOMLPLKVOIQYLDG--KYIDAKLK 1074
2015 KDAV--NGAOVHHISNKSISIGN--TYVNPDSLTNNIGTGKNNINDAISKYDEKVT 2071
1075 NN-NTVTSYNNFGA-----LPSWVVPYTAGSTIGLT 1105
2072 NGVNDLTOKGLNFGANDOKTQKAVHRKLGDTINTV 2108

RESULT 4
US-08-968-685A-10
Sequence 10, Application US/08968685A
Patent No. 6214981
GENERAL INFORMATION:
APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSLA, LAURA
TITLE OF INVENTION: MORAYELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York

COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968, 685A
FILING DATE: NO. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 869-8090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 2123 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-968-685A-10

Query Match 3.2%; Score 205; DB 3; Length 2123;
Best Local Similarity 19.8%; Pred. No. 1.7e-05;
Matches 286; Conservative 157; Mismatches 502; Indels 498; Gaps 72;

24 SASFGKSDKSNNTOLVNOARTL-----DANSVRLAGLGN-----GSLFNTVLVD 71
482 AATVG--OLKRYNQTAEALQTFYKVDKNGANDANSKIITVGNKKRPGDQVNT----- 535
72 VDDNFTIANGIIILKFTKPLVGLDISDCCGKYKQVYSYTTSRNRFDDQRQRAY 131
536 -----LKL-----KGNGVDVTTETNG-----TVTF 556
132 ALLVNDENVHLKRIINTSNRIGNNNNSKFEVIG-----GVNDPAHVIR 175
557 GINONNGILVNSTLNNGLSVKNNSKQLOVGADGITFTDISNKRGAIEV---TTR 613
176 FTDDGTKEFNTQTOGEIYNDILDA---PIPKDLHPDWYNYLQKRIIPNDVFAVVP 232
614 ITRFDG--IGFANNVGS-----LDANKPRLPPTGINSNGKEL-----TVQSAINP 656
233 WPGVRYSGTNADGMPDCNGQIT--NTDPIAOTKTTTPD-----NONPSTFNSGAMPANR 287
657 -----ATNG--GOLDPMNRUSTANTEKSSAATIKDLNLSQVPLTFAGDTGPNVTK 707
288 Y-----DSOLNVKHKIKTSFOLDEKRIINTSNRIGNR- 765
708 LGEILKYKGGKTAADDLTKNNIGVADSTDNSILTVK--IAKTLSDLD--AVNTKTLTLASDKV 765
319 -----NNNSKF-----VIGVDNPAHVIRFTD-----DGTREN 346
766 TVDSGNNTAKLONGDLTFSKONTGATPATNSKIITYVGD-----LKTQDNGIALDGTTY- 820
347 FTNTOGEIYNDILD--APILPKDLHPDWYNYLQKRIIPNDVNT---AVVPMVGVV 400
821 ITRKRVGFAKQDGLDSKPYLDK-----KIKVEVEITNGINAGGKATYGLSNITL 874
401 SGTNADG-----MFDGNGQITVNTD--PIAOTKTTTNDONPST 437
875 DATNATGHTVQLGIVSDTKTRAASIGDVLNMGFNLKNNGDAKDVSYTYDVT----- 928
438 FNSGAMPANRRYSOLNVKHKIKTSFOLDEKFFYPEMTGSEENKNITRLATGSLPSNER 497
929 FINGNATTAKYTTQKAS-----KVAIDVNVVDGTIHLTGADGNKNKNOIGKTYTLTKYT- 981
498 TWIIDIPGTPOVTLK-----EDSVNVFSRLYLSVNSLSFGDSIYIEGTSLESLWYVSF 553


```
Db 982 ----DAKGDAINFVSNGDDKALINAKDIADNLNTLA--GEIRMTKGTAD----- 1026
OY 554 PTRLSDLTALNOVKRDDIDIASSTDNCTTNGTTTADTSSGSGAGTGTTNTSQTVP 613
Db 1027 -----TALQFQVKVKV-----ENGDDN---DADTIT-----VGDATKTNQ----- 1060
OY 614 TLNTYRSFGIDSKPTSAKIDETNMADPVIEARIYAERLIGNEIPTNNGNFRTNI 673
Db 1061 -VNTLKLKGRKNGLDIOTNR-----DQTVTFGITWQSLKRGNTTILNNGLSIKRNA 1111
OY 674 G-----GVGFTSTGSRVYLRLASVNGDORPTNFOPELVYFVGLGYQQTGTG--FW 722
Db 1112 GNEQIOVGADGVKFAVNNGVV--GAGIDGTTTRITRDEIGFAGTNGSLDKSKPLSKDGIN 1170
OY 723 YGTYKLNNSPYDVLDSPRVGTETNFRRTSLTYPMGYL--TE-----EGARSS 772
Db 1171 AGGKKTIN-----IQSGEIAONSND-----AVTGKIKYDLKTELENKISSTAKTAQ 1216
OY 773 NTPYIRAGDTPESRSIFQSGYSDNTYEXIQSVLGF--DGIRNNLN-----VGKASSF 824
Db 1217 NSLHERSVADDEGONNTTVPYSSYPTSKTSVDYITFAGENGITTKVKNKGVVRYGIDQITG 1276
OY 825 LMSNR---BNPNGLEMIATYTLRSQIGLARTSGLPNQOPFGTHQVISVSPGDQPSIK 881
Db 1277 LTPKLTGVNNGNGKGV-----IDSQNGQNTITGLSN-----TLA 1311
OY 882 NRTTFPGNQLMFLFTNENKSSVYTLR---LADSSNPDDSSSPSLDIDVNEIGYI 937
Db 1312 NY-----TNDKGSVYRTTEOGKIKIDEDKTRAA-----SIYDVLGAFN 1349
OY 938 L-----PLLD--NSFYVNAAGNVALFSSNPGSGSYTAVNTFNQNLSDIAFEGSGAK--- 988
Db 1350 LOGNGAVVPFSTYTDVNFADGNATTAKYVDDTSKTSKYVDVNVDDTTIEKKDKLGY 1409
OY 989 ----YTSDFWGTIOFNPDEYLIQNGFT-----SQVANFYTNQSEFLNSLVDFTPANAG 1037
Db 1410 KTTTLTSTGTGA-----NKFALSNQATGALVKAISDVAHLNLTSGDIDGTAKAGASQANS 1464
OY 1038 TYRRVYVDDGNLTNQNLPKQIQYLDGKYIDAK-----LKNNNLVYFSYNNFG 1087
Db 1465 AG---VDADGN-----KYIYDSTDKYKQAKDGTVDKTEKAKDKLVAQAQTPDG 1513
OY 1088 ALPSWVYPTAIGS-----KVIYDSTDKYKQAKDGTVDKTEKAKDKLVAQAQTPDG 1511
Db 1514 TLAQMVKSVINKEQVNDANKKOGINEDNAFVKGLEKASDKTKNAATVGLNAV--- 1570
OY 1112 GLAIGPLRAQRKQ---DKGRTTFKKV--DTLTAAGSVYKKIITQ-----T 1155
Db 1571 -----AQTPLETFAGDTG--TTAKKLGELTIRKGGQDTNKLTDNNIGVAGTIDGFT 1619
OY 1156 ANVKKKPAALGAGKSG---DKKRAAARPAAPAKSAPKASSAPKTPGPK---SGAP 1207
Db 1620 VKLAKDLTMSVNAAGTKIDKGVSVDSGQAKANTPVLISANGLIDGKVISNVKGT 1679
OY 1208 KPT 1210
Db 1680 KDT 1682
```

```
RESULT 5
US-08-169-927-2
Sequence 2, Application US/08169927
Patent No. 5783441
GENERAL INFORMATION:
APPLICANT: Carl, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Mei
APPLICANT: Dasch, Gregory A
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ. ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1612 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-927-2

Query Match 3.1%; Score 199; DB 1; Length 1612;
Best Local Similarity 17.5%; Pred. No. 3e-05;
Matches 228; Conservative 170; Mismatches 390; Indels 516; Gaps 59;

OY 14 GGLAV-----FGALGSAFSGFKOS-----DKSNDTQLYNQARTIDANSVRLAG 59
Db 127 GGAIAINANDLSGSGITFAVCPSEVLEFNILNPINSRSSHYLNSKNRYN-----G 177
OY 60 QNGSLFNVLRDDVDNFITAANGTIIKLSFTKPLKGLDSDCGGKYKQIVSDYTSR 119
Db 178 GNCIL-----NTNCFIQVDSOFTFAGIKTIN-----IDDCG----- 209
OY 120 NRPDQROTAYYALVN---DEANVHLKRINTNSNRIGNRNNSKRVIGVDNPAHVIRF 176
Db 210 -----LMFNSTPDAAN-----TLNLQAGWNTIN-----FNGIDGFKLVLY 245
OY 177 TDDG--TKFNTFNOTGGEIYNDPILDAPILPKRLHPDWYNLYIQRLILPVDVTAVPMP 234
Db 246 SKNGAATERNVCTGCGNL-----KGI--ELMTAAVACK 278
OY 235 V-----GRVSG--TNADG----- 246
Db 279 LISLGAANAAYICTDNGAGRAAGFIYSVNGNATISGOVYAKNMVYQSNAGGYTFEH 338
OY 247 MEDCGNGQITNDPILAQTKTTTDNQNPSTFNSGAMPKANRRDSQNLNVKHR--IKTSFOL 304
Db 339 IVDVGLGCTTNEFTADSKVILITENSNGSYNFGNL-----DTQILVDPDKILKGNFTG 391
OY 305 DERINTNSNRIGNRNNSKRVIGVD-----NPAHV--IRFTPD 341
Db 392 DVKNNNGTAVGITFENANGALVASTDPNTIAVTINAIIEAGACVVELSGIHIDELGNG 451
OY 342 GTKFNFTNOT-----QGEIYVDEFLDA-----PILPKDLHPDWYNLYIQRLILP 386
Db 452 GSIFKLADGTVINCPVQNMNMNNMNLAAAGSIDLDSALITIGDIGNGVNAALQHTTLAN 511
OY 387 DVNTAVPMPVGVGSGTNADCGHFCGNG--OITPTD-----PIAQK 427
Db 512 DA-SKILALDGANIIGANVGAIHFQANGTITKLTFQONNIIVNFPDLIDTDTKTVVDAS 570
```



```

Db      911 TINTFVVGGLFDKKGNSN-----ISIAKGARFDDIDSKNLSITNTSSRYRTI 111
QY      887 FPGNOJMYLELTNEN-----NKSSVYTLRL-----911
Db      961 ISGN-----ITRKNDDLNTNEGSDTEMOIGDVSQKEGNTLTISSDKINIKQITIKAG 1014
QY      912 ADSNNDASSFSPTLIDVNEIGVILPLIDNSF-----YVNAAGVNALESSNPGSGSY 967
Db      1015 VDGNSDSDATNNANLTITKTELKLTQDNLISGFNKAETAKGSDLTIGNTN-SAGGTN 1073
QY      968 TAVTFNQNLSIDIAFEGSGAKYSDPFGTIOQRPDEYLIONGFTSQARNFVYNOSPLNS 1027
Db      1074 AKKVFENQ-VKDSKISADGKHVT-----LHS 1098
QY      1028 LVDETPANAGTNRVVVDDGNTLNQNLPLKYOIYLDGKYDAK--LKNNNLVTFESYNN 1085
Db      1099 KVTSGSSNNTE-----DSDNNAGLFI-----DAKNVTYNNNITSHKAVS 1139
QY      1086 FGLAPSMVYF---TAIGSTIGILAIMIIGLAGIPLRAQRKLODKGFTTKKVDTLTA 1142
Db      1140 ISATSEITTKTGTNNATTGAVNEITAGTGSILG-----GIESSSGSV-TLTA 1186
QY      1143 -----AVGSYKKIITOTAN-----VKKRPAALGAGKSGD 1172
Db      1187 TEGALVANSISGNTVYVANSALTLTLAGSTIKGTESVTTSSQSGD 1232

```

```

RESULT 7
US-08-480-604A-6
; Sequence 6, Application US/08480604A
; Patent No. 5736139

```

```

; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-604A-6

Query Match      3.0%; Score 190; DB 1; Length 2710;
Best Local Similarity 19.3%; Pred. No. 0.00029;
Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;

QY      42 VNQARTLDANSVRLAGLGNGSLFNVLVDVDDNFITPAANGTIKIDSTFKPL-----94
Db      1071 INMSLSIAATVASIVIGAEVTFILPLIAGISGIPSLVNNELIHDKATSVNVYFHLIS 1130
QY      95 -----YGLSDDCGKVKVQIVSDYTTSSNRREDQKQTRAYVALVNDANVHLKRIINTN 149
Db      1131 ESKRYGPKLTEDD-----KILVPIDDLVISEIDF-----NNN 1162
QY      150 SNRIGNRN-----NNSKFVIGGVDPNPAHVIRFTDGTREFTNTOGEIVNDFILDAP 203
Db      1163 STKIGTCNLIAMEGSGSHVYTGNDH-----FFSSPSISSHPLSLISYSA--1207
QY      204 LPKDLHPDWNYNLVIOKRI--LPNDVNTAVVWPVPGRY-----SGTNADGMPGCGNG 253
Db      1208 -----IGIETENLDFSKRIMLPN-APSRVFMWETGAVGLRSLDNDGTRLLDSIRDLYPG 1262
QY      254 Q-----ITNDPIAO-----TKTTDQNSSTNSGAMP 283
Db      1263 KFYRWRYAPFDVATITLKPVEYEDTNIRIKIKDKOTRNFIMPTITTNELRNLSY---SEDG 1319
QY      284 ANNRYSQNLNVKRIKTSFQDLDERINTNSRIGNRNNNSKFVIGGVDPNPAHVIRFTDGT 343
Db      1320 AGGTYSILL-----SSYPSISNINLSKQDWM-----IFNDINEVRELSIENGTL 1363
QY      344 KFNFTNTOGEIVNDFILDAPILPKDLHPDWNYNLVIOKRI--LPNDVNTAVVWPVPGRYSGT 403
Db      1364 K-----KGLIKD-----VLSKI---DINKNKL-----ITGN 1387
QY      404 NADGMPGCGNGGQITNDPIAOQTKTTDNDONPSFTN-----SGAMPGANRNYDSQL-454
Db      1388 QTID-----FSGDIDNDKRYIFLTCELDIDISLIIEINLVAKSYLLSDDKNYLLSNLS 1442
QY      455 NVKRIKTSFQDLDEKFEYYPWMTGSEENKNTRLATGSLPSENERVWILDIGTPQVTLKED 514
Db      1443 NTEIEKINT-LGDSKNINAVYTTDESNNKFGAISKTSQSKIIHY-----KRD 1488
QY      515 SVNVFSRLYNSV---NSLSFIGDSIYIFCTSELPSL---WYSFPTRLS-----DITAL 563
Db      1489 SKNIL-EFYNDSTLEFNSKDFIAEDINVFEMKDDINTTGYVYDNNNDKSIDPSISLVS 1547
QY      564 NOVKTDDIEASSTDNGTTNGTTTATPSSGSGAGAGNTNTSQVSNFTPLMYRBSFGI 623
Db      1548 NOVXV-----NGLYLNEYSYSTLDFVKKNSDGHNTSNFENFLD--NISFWKLGF 1597
QY      624 DSKPTSAK-----IDETNADDPNVEARIVAEYRLGIONPIPTNAGFNTRNTIGVGFT 679
Db      1598 ENINFEVDKFTLVGKNTLG-----YVEFICDNKNKNDI-YFEWMTSSSKRTIFS 1647
QY      680 STGSRVVLASVNGDORPTG-----NFQPLVYFGVLGVOQTGTGF-----721
Db      1648 GNGRNVVPEIYND---TGEDISTSLDFSEPLYGIDRYINKVLIAPIADLYTSLININTN 1704

```

	REFERENCE/DOCKET NUMBER:	OPHD-01308
	TELECOMMUNICATION INFORMATION:	
	TELEPHONE: (415) 705-8410	
	TELEFAX: (415) 397-8338	
	INFORMATION FOR SEQ ID NO: 6:	
	SEQUENCE CHARACTERISTICS:	
	LENGTH: 2710 amino acids	
	-type: amino acid	
	TOPOLOGY: linear	
	MOLECULE TYPE: protein	
	US-08-405-496A-6	
	Query Match	3.0%; Score 190; DB 2; Length 2710;
	Best Local Similarity	19.3%; Pred. No. 0.00029;
	Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;	
QY	42	INQARTDANSVRLAGLGGSLFNTVLARDVDNFTTANGTIKIDSTFKL-----94
DB	1071	INMSISIAIVASIVLGIAEVTIPLRLIAGISAGISPLVNNELIHDKATSVVNFNHLIS 1130
QY	95	---YG-LDLSDDCGGKRVKQIVSDYTSNRRFDQRTAAYVALLVNDEANVHLKRITNN 149
DB	1131	ESKKYGPLKTEDD---KILVPIDDLYISELDF-----NNN 1162
QY	150	SNRIGNRN-----NNSKFVIGVDNPAHVIRFTDDGKTFNFNQGEIVNDFILDAPI 203
DB	1163	SIKLTKTCNLMIEGSGHVTGNIDH-----PFSSPSISSHIPSLSIYSIA-- 1207
QY	204	LKRDLPWNYNLYIORKI--LPNDVNFVPMVPGRV-----SGTNADGMPDGGNG 253
DB	1208	---IGLEINLDFESKKIIMLPN-APSKVPMWETGAVGLRSLSENDGTRILDSIRDLYPG 1262
QY	254	Q-----TNTDPIAO-----PKTTDNONSTFNSGAMPG 283
DB	1263	KFYRFYAFEDYAIITLKEVYEDINIKIKLDKTRNFMPTITTEINRKLSTY---SFDG 1319
QY	284	ANRRYSOQLNVKHKIKTSPOLDERINTNSNRIGNNNNSKFVIGVDNPAHVIRFTDDGT 343
DB	1320	AGGTYSLL-----SSYIPSTINILSKDMLL-----IFINDEBREISIEGNTI 1363
QY	344	KFNFTNTOGEIVNDFILDAPIPKDLHPDWNYNLYIORKILPNDVNFVPMVPGRVSGT 403
DB	1364	K-----KGLIKLD-----VLSKI--DINKKKL-----IGN 1387
QY	404	NADGMPDGGNGOITMTDPIAQTKTTDNONSTFN-----SGAMPGANRRYSOL- 454
DB	1388	QTID-----FSGDIKDKRATFYLCELDKISILTEINLVAKSYSILSGDKNYILSMLS 1442
QY	455	NVKHKIKTSPOLDERKRYFPMWETSEENKNTITRLATGSLPSNRMYLLIDPGRVYLTED 514
DB	1443	NTEIKINT-IGLDSKNAIYWTBESNNKRYGALISKTSQKSIHH-----RKD 1488
QY	515	SVNFESRLYINSV---NSLSFIDSIYFEGTSELPRL---WYSPFTRLS-----DVAL 563
DB	1489	SKNLI-BEYNDSTLEFNSKDFIEDIENVFEMKDDIINTGKYVDNNTDKSIDPSISLYSK 1544
QY	564	NOVKTIDIEASTDNGTNTGTTATTATTSAGTSAGAGTNTNSQTSVNPNTINTYRSQGI 623
DB	1548	NOVKY-----NGLYNFSYSSILDEYKNSDGHNNNSPNNMLFLD-NISFWKLQEP 1597
QY	624	DSKPTYSANK---IDETNMADPNVIEARITAEYGLQINPIPTTNGNFRNTIGGVGTF 679
DB	1598	EMINIVIDKRYFLVAGKTMIG-----YVEFICDNKKNIDI-YREQEMKTSKSTIIS 1644
QY	680	STGSKRVILASVNGQORPTG-----NFQPLVLYFGVLYGQOQRTGTF----- 721
DB	1648	GNGRNVVVEPIYNPD---TGEDISTSDOFSEVPELYGIDRYINKVLAPLDYLSLININTN 170
QY	722	---WYGTKLNNSPYDV-----LDSP-----RVQGNENQFRSTLSLYRPMVGUYLDEEG 767
DB	1705	YYSNEYPEPIIYLVNPTPEHKKKNINILDSSEFEYKMTSEGSDF-----ILVRYLEE-- 1757
QY	768	ANSESTNPIIRAGQDPPEARSIFGSGYSDNTYVEYIQSVLFGDGIIRNNLNVGKASSPLNS 827

Db 1755 -----SNKKIL-----OKIRIGILISNTQSPNMSIDFDIK-KLSIGYIMSNFKSF 1800
QY 828 NRPNP-----NGLEMAATTYLRSQ-----IGLARTSGL-----PNOQPFQTHHOVIS-- 870
Db 1801 NSENELDRDLKFKIDNKTYYYDEDSKLVKGLININNSLFEYFDPIEFELVJGMQITNK 1860
QY 871 -----VSPGDQSSIKINRTIPFGNQLWYELFTNENKSSVYTLRLADSDNPASSPSP 925
Db 1861 KYEFDINTGALITSK-----LTKGK-----HFFRNNDG-----VMOLGVFKPGDFEYFAP 1907
QY 926 TSLIDVNEIGVILPLIDNSFYVNAAGNALFSSNPGSYAVNT-----FQNLSDIA 981
Db 1908 ANTQNNNIEGQAI-VYQSKFELTN--GKKYFDNNSKAVTGMRILINNEKYYFNP-NATA 1963
QY 982 FEG-----SGAKY-----TSDFMGTIO-----FKPEYLIQNGFTSQVARNFTVNO-- 1022
Db 1964 AVGLQYIDNNKYFFNEDTAIIISKGMQTVNGSRYYEPTDPAIAFNGYKTIIDGKHFFYEDSDC 2023
QY 1023 -----SFLNSLVDETPANAGNYRVVDDPGNLNOMLPLKVOIQYIDGK--YYDAKL.1073
Db 2024 VKIKGVSTSNGEFEFAPANTYNN-----NIEGQAIYQSKFELTNKGYTFD--- 2071
QY 1074 KNNNLYT-----FSYNNFGALPSM 1092
Db 2072 NNSKAVTGLQITIDSKKYYENTNTAEATGM 2101

RESULT 9
US-08-915-136-6
; Sequence 6, Application US/08915136
; Patent No. 6230960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPND-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-915-136-6
Query Match 3.0%; Score 190; DB 3; Length 2710;
Best Local Similarity 19.3%; Pred. No. 0.00029;
Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;
QY 42 VNQARTLDANSVRLAGLGGNGSEFNTVLRVDVNDNFITAAAGTIKIDSTFKPL----- 94
Db 1071 IMMSLSIAAFVAVSIGAEVETIFLLPIAGISAGIPSLVNNELIHDKATSVYVNPENHLS 1130
QY 95 -----VG-LDSDDCGGYKKQIYSDYTSRNRPDQGTARAYVALVNDENAVHLKRINTN 149
Db 1131 ESKKIGPLKTEDD-----KLVPIDDLVSEIDF-----NNN 1162
QY 150 SNRIGNRN-----NNSKEFVIGVDNPAHVIRETDDGFENFTNOTGELVNDPILDAPI 203
Db 1163 SIRKLCNILLAMEGGSGHVTGNIDH-----FFSSPSISSHIPSLISYSA-- 1207
QY 204 LPRDLHPDYNLYIQRTI--LPNDVNTAVVPMVGRV-----SGTNADGMEFCGNG 253
Db 1208 -----IGIETENLDFSKIMLPLN-APSRVFMETGAVGLRSLNDGTRILLDIRLYG 1262
QY 254 Q-----ITNTDPIAQ-----TKTTDNONPSTFGSAMPG 283
Db 1263 KFYWRFAFFDAITTLKPYVEDTNKIKLKDTRNFIMPITTTNEIRKLST---SFDG 1319
QY 284 ANNRYDSQNLVKHRIKTSFOLDERINTNSRIGNRNNNSKEFVIGVDNPAHVIRETDDGT 343
Db 1320 AGCTYSILL-----SSYPISPINIMLSKDDLM-----IFINDNEVEISEINGTI 1363
QY 344 KENFTNOTGELVNDPILDAPIPKDLHPDYNLYIQRTILPNDVNTAVVPMVGRVSGT 403
Db 1364 K-----KGKLIKD-----VLSKI--DINKRKL-----IIGN 1387
QY 404 NADGMEFCGNGOITNTDPIAQTKTTDNONPSTFN-----SGAMPGANRRYDSOL- 454
Db 1388 QITD-----FSGDIDDKDKRIYFLTCGLDCKISLIIENLVAKYSLLSDGKRYLISNLS 1442
QY 455 NVKHKIKTSFOLDERKFEVPEWGTSEENKNITRLATGSLPSNEREYWLIDIPGTQVYLKED 514
Db 1443 NTEKINT--LGDSKNIAVYTDENSNKKFGALSKTSQSKSIHY-----KKD 1488
QY 515 SVAVFESRLYNSV---NSLSFGDSIYIRGTSLSPL---WYISFPTRLS-----DITAL 563
Db 1489 SKNII--EFYNDSTLERNKDFIEDINVEFKDDINTYTKYVYDNTDMSIDPSISLSYK 1547
QY 564 NOVKTDDIEASSTDNCGTTTGTGTTTADTSGSGAGAGNTNTSQTYSNLTLYRSFGI 623
Db 1548 NOVKV-----NGLIYNESYSSSLDPLVKKSDGHHNTSNMNLFLD-NISFMKILGCF 1597
QY 624 DSKPTSANK-----IDETNADPVPVIEARIYAEVRLGIONEIPITNAGNFRNTIGVGFT 679
Db 1598 ENINFPYIDKYFLVKTGNTG-----YVEEICNNKNKIDI-YGEMKTSSSKSTIFS 1647
QY 680 STGSRYVLASTYNGDORPTG-----NFQPLVYFGVLGIGQRTGTG----- 721
Db 1648 GNGRNVVVEPIYND--TGEDISTSLDFSEPLVGLIDRYINRVLLAPDLTYSLININTN 1704

RESULT 10
US-08-957-310-6
Sequence 6, Application US/08957310
Patent No. 6365158
GENERAL INFORMATION:
APPLICANT: Williams, James A.
APPLICANT: Kirk, John A.
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
TITLE OF INVENTION: DISEASE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,310
FILING DATE: 23-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 24-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.

Query Match	13.3%	Score 190;	DB 4;	Length 2110;	
Best Local Similarity	19.3%	Pred. No. 0.00029;			
Matches 238;	Conservative 155;	Mismatches 459;	Indels 378;	Gaps 61;	
QY	42 VNCARLIDANSVR	LAGLGGNSLFNV	LDVYDNFT	TANGIIILDSFTPL-----94	
DB	1071 INMLSIAA	VASIVGIGAEV	TEFLPL	IGISAGISPLVNNLILHDKATSVNVFNHLS 1130	
QY	95 ----YG-L	LDSDCGKVKQI	VSDDYTS	SNRPEDQRTAYAYALLVNDENVHLKRITFN 149	
DB	1131 ESKRGG	PLKTEDD----	KILVPI	DDLVYSIDF-----NNN 1152	
QY	150 SNRI	GNRN-----	NNSKFV	IGVDNPAHIRTDOGTKEFNTOGELVDFLLDAPL 203	
DB	1163 SIKGTG	CIILAMEGSG	HHVTGNIDH-----	FFSSPSISHPISLSTISA-- 1207	
QY	204 LPKDLHP	WMYMLYIQRK	--LPNDV	TAVVPWVGRV-----SGTNADGMDGCG 253	
DB	1208 ----	IGITENL	DFSFKIMLPLN	-APSRVPMETGAVGLRSLNDGTLLDSIRDLVPG 1262	
QY	254 Q-----	ITNDDPL	IAO-----	TKTTNDONPSTFNSGAMP 283	
DB	1263 KYWMEFY	AFEDYAIT	TLTKPEY	EDNIIKKDKDTRNIMPTVTINIRKLSTY---SFDG 1319	
QY	284 ANNRYSOL	NKHKRITS	FQDLER	INTNSRIGRNNNSKFVIGVDNPAHYRTDDGT 343	
DB	1320 AGGT	SLL-----	SSYPI	STININSKDLM-----LPNDNEVRELSIENGCI 1363	
QY	344 KKNFN	NOGELV	NDFIL	DAPILPKDLHPWMLYIQRKLLPNDVNTAVVPWVGRVSGT 403	
DB	1364 K-----	KGKLIK	D-----	VLSKI--DINKNKI-----TIGN 1387	
QY	404 NADG	MEGDCG	NOITN	DPIDPAOTKTTTNDONPSTFN-----SGAMPANNRYSOL- 454	
DB	1388 QTID	-----	FSDID	DKDRYIFLTCGLDDKLSLIIENLVAKSVLSLSDKHYLLSNLS 1442	
QY	455 NVKHKR	ITSFQDL	EKFEVPE	WTSGSEKNKTRILATGSLPSNERKWIIDIGTQVYLKED 514	
DB	1443 NTLEK	INTLGD	SKNIAV	YDESNNKYFGAISTOSKSIHH-----KRD 1488	
QY	515 SVAV	FRKLNSV	-----	NSLSFGDSIYIEGTSELPSL---WYSEFTRL-----DIAL 563	
DB	1489 SKNLL	-EFYD	SLERNSK	DFLAEIDNVMKODINTIGKYYVNNNDKSIDPESISLVSK 1547	
QY	564 NOVKT	DI	EASST	DNCTTNGTTTAAIDSSGSGTACGTGNTNTSQTYSNPLNTYRSQFI 623	
DB	1548 NOVXV	-----	NGYL	INESYSSYLDFVKNSDGHHNTSFMFLFD-NISMKPLGEF 1597	
QY	624 DSKRT	SANK-----	IDET	NADPVI	EARIYAEVRLGIONEIPITNAGFINNTIGVGVOFT 679
DB	1598 ENIN	FEV	IDKFT	FLVGRKNG-----YVEFIC	NNKNDI-YFGEMKTSSSKSTIFS 1647
QY	680 STGS	RVVLR	ASVND	QRPFG-----NEQPELV	VEGLYQOQTRGTG-----721
DB	1648 GNGR	NNV	VEEPI	YNDP-----TGEDI	STLSDFEFLYIDKRIKVVLAADLVLSLINNTN 1704
QY	722 ----	WYGT	YILNNS	PEYD-----LDS	---RVGETNOFRRTSLIYPWVGGLTEBEG 767
DB	1705 YYSNE	YELLV	MPNPF	HHKVNINL	DDSSSEFEKMTSGDSF-----LIVRYLEE-- 1754

QY 768 ARSFNTPIYRAOGDPESRSIFOSGSDNTYIEIOSVLGFDGIRNNLANVKAASPLNS 827
 Db 1755 -----SNKKIL-----OKIRIKGLSTFOSFNKMSIDFKRIK-KLSLGLTMSNFKSF 1800
 QY 828 NRNP-----NGLEMTAATYLRQ-----IGLARTSGL-----PNOQPEGTHOYIS- 870
 Db 1801 NSENELDRLHGLKIIDNKTYYYDEDSKLKVLGININSLEFDPLEFNLVGTQMOTNGK 1860
 QY 871 -----VSPGPFSSINIRTIIFPGNOLMYFLFTNENKSSVYLRRLADSSNPQASSFSF 925
 Db 1861 KYIFEDINTGALTSYK-----INGK-----HEFYNNDD-----VMQIGVEKFGPGFEYFAP 1907
 QY 926 TSLIDVNEIGVILPLDINSFYTVNAGNVALFSSNPGSPSYTAVT-----FQNLSDIA 981
 Db 1908 ANQNNNIEGOAL-VYOSKFLTLN--CKKYFDNNSKAVYGMRLINNEKTYFNPB-NATA 1963
 QY 982 FEG-----SGAKY-----TSDFWGTIO-----FKPDEXLIONGFTSOVAREFVYNO- 1022
 Db 1964 AVGLQVLDNNKYFENPDTALISKGWOTVNGSRYYFDIDFAIAFNGYKTIKDGKHFYFSDC 2023
 QY 1023 -----SFLNSLVDETPANAGTNRVYVDPDGNLTNQLPLKVOIOLYLDG--YYDAKL 1073
 Db 2024 VAKIGFSTNSNGEYFAPANTYNN-----NIBGOALVYOSKFLTLNGKKYYED- 2071
 QY 1074 KNNNLVY-----FSYNNFGALPSW 1092
 Db 2072 NSKAVYTGLOTIDSKKYFNTNTAEATGW 2101

RESULT 11
 US-10-011-366-6
 ; Sequence 6, Application us/10011366
 ; Patent No. 6573003
 ; GENERAL INFORMATION:

APPLICANT: Williams, James A.
 Link, John A.

TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES
 OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE
 DISEASE

NUMBER OF SEQUENCES: 22
 CORRESPONDENCE ADDRESSES:

ADDRESSEE: Medlen & Carroll
 STREET: 220 Montgomery Street, Suite 2200
 City: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/011,366

FILING DATE: 16-No. 6573003-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/957,310

FILING DATE: 23-OCT-1997

APPLICATION NUMBER: US 08/329,154

FILING DATE: 24-OCT-1994

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01121

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338
 ; INFORMATION FOR SEQ. ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2710 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
 US-10-011-366-6

Query Match 3.0%; Score 190; DB 4; Length 2710;
 Best Local Similarity 19.3%; Pred. No. 0.00029;
 Matches 238; Conservative 155; Mismatches 459; Indels 378; Gaps 61;

QY 42 VNQRITLANSVYLAGIGGSLFENLYLRVDNFTIANGTIIKIDSFKPL----- 94
 Db 1071 INMSLSIATVAVSIGAEVTLFLPLIAGISAGISLVNNEILHDKATSVNYFNHLS 1130
 QY 95 -----YGLDSDCGGYKQIYSDYTSRNRFDOROTRAYVALVDEANVHLKRIINTN 149
 Db 1131 ESKKYGELKTEDD-----KILVPIDDLVSIEDF-----NNN 1162
 QY 150 SNRIGNRN-----NNSKEVIGVDNPAHAVIRFTDGTCKFNFTNOGELVNDPILDAPI 203
 Db 1163 SIRLGTGCIILAMGSGSGHTVTGNIDH-----FFSSPSISHSIPSLISYSA- 1207
 QY 204 LPKDLHPDMYNLYIQRKI--LPNDVNTAVVPPWPGVRY-----SGTNADCGMFDGNG 253
 Db 1208 -----IGIETENLDPSKKIMLPN-APSRVEMETGAVGLRSLSENDGTRILDSTRIDLYPG 1262
 QY 254 Q-----ITNTDPIAQ-----TKTTDQNPSTFNSGAMP 283
 Db 1263 KFYWRFYAFEDYAITTLAKPYEDNINIKIKDKDQTRNIMFTITNELRNLSY---SFDG 1319
 QY 284 ANNRYDSQLNVRKIRIKTSFQDLDERINTNSRIGNRNNNSKFVIGVDNPAHAVIREFTDGT 343
 Db 1320 AGGYSLLL-----SSYPISITNINLSKDLW-----IFNIDNEVEIREISENGCTI 1363
 QY 344 KFNFTNOGELVNDPILDAPILPKDLHPDMYNLYIQRKI--LPNDVNTAVVPPWPGVRYSGT 403
 Db 1364 K-----KGLKLD-----VLSKI--DINKNKL-----ITGN 1387
 QY 404 NADDMEDCGNGOITNTDPIAQRTKTTDQNPSTFN-----SGAMPANNRYDSQL- 454
 Db 1388 QTID-----FSGDIDNDRYIFLTCEDDKISLIEINLVAKSLSLLSDGKNLISNLS 1442
 QY 455 NVKRIKTSFQDLDEKPYRPMYTGSEENKNTIRLATGSLPSNERMYILDIGTPQVTLKED 514
 Db 1443 NTEIKINT-LGLDSKNIAVNYTDESNNKRYGALSISKTSOKSIINHY-----KRD 1488
 QY 515 SVNVFSLYLSNV---NSLSFIDGSIYIEGTSELPRL---WYVSFPRLS-----DTAL 563
 Db 1489 SKNLT-EFYNDSTLEFNSKDFIAEDIVFEMKDDINTTGGYYVDNNDKIDISISLSVK 1547
 QY 564 NOYKTDIEASSPDGTTNGTTTADTSSGAGTGGTNTTSSQVTSNPLNTLYRSFGI 623
 Db 1548 NOYKV-----NGILYNESYSSLYLDEYKNSDGHNTSNFNLELD-NISWKLFGF 1597
 QY 624 DSKPTSAK---IDETNMAPVNIPEARIAEYRGLIGNEIPIITNAGNFTIRNTIGVGFT 679
 Db 1598 ENINFIIDKYFTLLGKTNLG-----YVEFICDNKNKNDI-YFGEWMTSSSKSYIFS 1647
 QY 680 STGSRVYLASYNDDORPTG-----NFOPFLYVGLYQOQTGTGF----- 721
 Db 1648 GNGRNVAVEPIYND--TGEDISTSLDFSEYEPYGLIDRYINKVLIAPDLTYSLININTN 1704
 QY 722 -----WYGTYKLILNSPYDV-----LDSP-----RVGTEYQFRSTSLTYVMGGLTEEG 767
 Db 1705 YSNSEYRPELIIIVLNPFTFKKVNININDSSSEFYKWSYEGSGDF-----ILVRYLDE-- 1754
 QY 768 ARSFNTPIYRAOGDPESRSIFOSGSDNTYIEIOSVLGFDGIRNNLANVKAASPLNS 827
 Db 1755 -----SNKKIL-----OKIRIKGLSTFOSFNKMSIDFKRIK-KLSLGLTMSNFKSF 1800

[illegible]

OY		221	-ILPN-	-VNAFVMPV-----GRSGINADDG---	PDCNGOITN--TDPAQFT	266
Db		1426	FALANDSKSVATSDKSLGTNGNKAVNTDGLTFARFARDSKTGDDANHLHGLASTLT		1485	
OY		267	TTONONSPTEPNSGAMGANNRRYSOLNKHRTKSFOLEDRINTNSNRIGNRNNK		323	
Db		1486	DTLNSGATTNLGGCITIDNEKKRAASVDVLNMGW----	NRGVKPAPASANNVOENID	1539	
OY		324	-----FVGIVDPNAVHTRFDODTKFPFNMQOGELVYNFILDABLPKLHPDM		374	
Db		1540	FVATYDVDFVSGDKTSTVESKKNCKREVEKIGAKTSYIKR-----		1583	
OY		375	YN--LYIQRLINDPVTAFAVPWPGRVSGTNADGMFCNGOITNDPLA		Q 425	
Db		1584	HNGLFGEKELKANNNGVVTETDGK-----DENGLVATAKAYDAVNKGWR		1632	
OY		426	TKTYYDN-QNS-----TF- NSGAMGANNRYDSOLNKHRTKS- FOLDEKF		470	
Db		1633	VKTGANQNDDFAVAAGTNNVFADONGTTAEVYKANKGSITYYXNVKVADGKLKDGDK		1692	
OY		471	VYPEWTGSEENKNITRLATGSLPS-----NERVIIDI--PGPIOV		509	
Db		1693	IYADJT-----VLTVAADGKVTPANNNGDKKEFVASGLADLANKLSMTARAEGEGEV		1745	
OY		510	TLKESVAVFERLLINSVNSLSF-IGDSIYEFTSELPSLMWYSFPRLSDLITALNOYKT		568	
Db		1746	--DPNASAQ-EVKAGDKYTFEAGBNLKI--KOSGDFTYSILKELKLDTS-----		1791	
OY		569	DIIEASSDNDG-----TNGTTTTADISGSGTAGCNTNTTSSQ-----		608	
Db		1792	--VEFKDANGCGESESTEKITRKDGLTTPANGAGAANTANTISYIKDISAGNKAVTNV		1849	
OY		609	-----TVSNPTL-----MYRSF-----GIDSKEPTSANKIDET-----		636	
Db		1850	VSLCAKFGGHLANGYADEFEKHYNAXKDLTLMDEKGDANNPTAADMTAAVGDGL		1909	
OY		637	NW-----ADPNVTEARTAEVYLGIQ--NEPIPMANCFIRNYIGGFSTGSRVY		686	
Db		1910	GWVISADKTTGBEN-----OEIAOVARNANEVKEPKSG-----NCINGSKTLNCTRYI		1957	
OY		687	LKASYNDGORPTGNFPOPLFYFGLYGOQ--TRIGTWYGXYKLNNSPDYDLSRVCYT		744	
Db		1958	TPELAKGEVYKSNEFT--VNADSGSEFINLVKKGDMTY-----SKEDIDPATSKPMNG		2007	
OY		745	ETNOFR-----RTSLTYPMVG-GYLT-EGARSFSNTPYLRACDTPERSSIF		790	
Db		2008	KTEKYAKENGRVVSANGSKTEVTLTKKSGGYVTQGVADATLAKSGFELGLADAEEAKAP		2067	
OY		791	QSQSYDN--TYEYIOSYLGFDGR--NNLNQGVKSSFLNSRPNGCLEMIATYLR		846	
Db		2068	AESAKKDQLSKDAKETVANADKVRFANGTNSAIYEST--DANDGVY--TTFEVT		2122	
OY		847	QIGLARSGPLNOOPFGTT--HOVISVSPGDOFSSIKNIRTIFFGNOLWFELTENNKK		903	
Db		2123	DV-----ELPINOITYMDANGKNKIYKADGK-----WELNADGTAS		2159	
OY		904	SSVYTRLADSSNPDASSFSPISLIIDVNEGILPELDNSFYTYNAAGNALFPSSNBGS		963	
Db		2160	NKEVTL-----GNVDANK--KVAVYTLNGA--DKWYTT-NADG-----		2193	
OY		964	PSYTAVMTFMNONLSIAFEBSGAKYTSDFWGTOIKRPDE-----YLIONFTSOVARN		1017	
Db		2194	----AADTKG-----EVSNDKSVSTDEKHVVRLDPNNSNGSKGVADIVANJEIS-		2239	
OY		1018	EYTNOSFLNSLDFTPANAAGTNNRYVVDPOGNTLNOQLPLKYOIOYLDGRYYDAKLKNN		1077	
Db		2240	-ATSDAINGSOLIVAVAGVTN-----LAGOVNLEKKVNVKG-KRAD		2280	
OY		1078	LVTFEYNNNGALPWWVPTAIGSTGLIALMILL--GLAIGTFLAOKRLDDKG		1129	
Db		2281	AGTASMLASOLPQAOTMP--GKSUVALTAGSSYOGONGLAIGV-----SRISDNG		2237	


```

US-09-669-974-33
; Sequence 33, Application US/09669974
; Patent No. 6333173
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/669,974
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-669-974-33

Query Match          3.0%; Score 189.5; DB 4; Length 2353;
Best Local Similarity 18.6%; Pred. No. 0.00025;
Matches 256; Conservative 177; Mismatches 479; Indels 463; Gaps 67;

OY 30 KQSKSDNTOQVNOATITDANSVRLAGGONGSLNPTVLRVDVDFITANGTIKIDS 89
DB 1141 KQSKR--DFYSLDDPIL--GLTITLGGTANGNDGTGY--NKDGILITLNGAAGTDA 1196
OY 90 FTKPLGLDSDDCGGYK---YQIVSDYTSRNRDQRPATYALLVN--- 136
DB 1197 SNGMTISYTKDGISAGKKEITNVKSLKTYKTONADEQDEFAAAYANANEVEVGK 1256
OY 137 -----DEANVHLKRIINTNSNRI----- 153
DB 1257 NGATVSAKTNDNGKHVTIDYAEKAKVDGLEKPTDGKIKKAVNTDGNLNLJVDATKAS 1316
OY 154 -----GNRNNSKFVIGVDNPAHVIRFTDDGKTFKNTNOGEIYN 195
DB 1317 VAKGEFNAVTTDATTAQGTANENERGKYVVG-----SNGATATETDKKKAATVG 1365
OY 196 DF---ILDAPILPRDLRD-----WYNLYIORK----- 220
DB 1366 DVAKAINDAATFEKENDDSATIDDSPTDGDANDALAGTLITLAKKRNILKVRDKNIT 1425
OY 221 -LLPND--VNTAVVPMV-----GRVSGTNADDGM--FDGNGQITN--TDPIAOTK 266
DB 1426 PALANDLSKATVSDKISLGNGKNKVNITSDTKGLNFAKSKTGDANIHILNGIASLT 1485
OY 267 TTDNONPSTFNSGAMPGANNNRYDSOLNKHRIKISFOLDEHINTNSNRIGNRNNSK-- 323
DB 1486 DFLNSGATTNLGGGIDNEKKRAASVYKVLNAGW-----NVRGKVPASANNOVENID 1539
OY 324 -----FVIGVDNPAHVIRFTDDGKTFKNTNOGEIYNDFILDAPIILPRDLRW 374
DB 1540 EVATYDTVDFVSGDKTTSVYTESKDNKRTEVKATGATSYIKD----- 1583
OY 375 YN--LYIQRKILPNDVNTAVVPMVGRVSGTNADDGMFDGNGQITNDPIA-----Q 425
DB 1584 HNGKLETKELKANNNGVTJETDGK-----DEGGLVTAARAVIDAVNKAQWR 1632
OY 426 TKTTTND--QNP-----TF--NSGAMPGANNNRYDSOLNKHRIKIS--FOLDK 470
DB 1633 VKTTGANGDNDPAFYASTNTTFADNGTITAEVKKANDGSTIVYKVKVADGKLDDGK 1692
OY 471 VPEWTGSEENKNITRIATGSLPS-----NERWILDI--PGTPOV 509
DB 1693 IVADDT-----VLIVADGKVYAPANNDGKRFVDASGLADALINKLSWTATAGKEGTGV 1745

```

```

OY 510 TLKEDSVNFSRLIYANSVNSLSF--IGDSIYIEGTSELPSLWYVSFPTRLDPLANQVKT 568
DB 1746 ----DPANSAG--EVKAGDKVFEKAGDNKLT---KQSKDFTYSLKKELNDLTS----- 1791
OY 569 DDEASTDNGT-----TTNGTTTADTSGSTGACTGNTTNSQ----- 608
DB 1792 --VEFKDANGTSGSESTKTKDKGLTITPANGAGAGANTANTISTYKGISAGKAVTNV 1849
OY 609 -----TVSNPTL-----NTRYSP-----GISKPTSAKIDET----- 636
DB 1850 VSGLKKFGDGHLANGTVADEFKHNDNAKDLTINDEKAGDNNPVAANTAAVGDRLGL 1909
OY 637 NW-----ADPNVTEARITAEYRLGIQ--NEIDITNAGNIRNTIGCVFTSGSVY 686
DB 1910 GWYISADKTTGEBN-----QEYNAOVANANEVEFKSG-----NGIVSGKTLNGTGYI 1957
OY 687 LRASYNGDORPTGNEPFLYVGYLGYOQ--TRTGFWYGYKLLNNSPYDLDSPRYGT 744
DB 1958 TPELAKGEVVKSNMEFT---VKNADSGEYNLVKVGDMY---SKEDIDPATSKPMTG 2007
OY 745 ETNOFR-----RTSLYYPVWG--GYLT--EGARSFNTPIYIRAGDTPESRSIF 790
DB 2008 KTEKYKVENGVYVANSKTEVTLTKKSGYVYGNQVADAIKSGFELGLADAAAEKAF 2067
OY 791 QSGYSDN--TXYIQSYVLCFDGIR--NNLNVGKASSFLNSRPNNGLEMTAATYLR 846
DB 2068 AESAKDKOLSKDKAEYVNAHDKVFANGLTKVSAATVEST---DANGKV--TTTFVKT 2122
OY 847 QIGLARTSGLPNOQPFCTT---HOVISVSPGDQSSIKNIRITFPGNOLMYLFTINENK 903
DB 2123 DV-----ELPDIQIYNTDANGKRIYKADGK-----WYELNDMGTS 2159
OY 904 SSVYTLRLADSSNDPSSSPTSLLDVNETGYLLPLDNEFYTVNAGVALLFSNPGS 963
DB 2160 NKEVYTL---GNVDANGK---KVRKYTEGA---DKWYTT--NADG----- 2193
OY 964 PGSTAVNTFNONSLDIAFEGSGAKYTSDFWGTIOFKPDE-----YLLONGTTSQYAKN 1017
DB 2194 ----AABKTKG-----EVSNDKVSSTDEKHVYRLDPNNSNGKGVYDINVANGIS-- 2239
OY 1018 FVTNOSFLNSLVDETPANAGTNRVYVDPDGNLNLQNLPLKYQIYLDGKYDAKILNNN 1077
DB 2240 -ATSDAINGQLAAVAGVTN-----LAGOVNNEEGKVNKVG--KRAD 2280
OY 1078 LVTFESYNNFGALPSPWVPTAIGSTGLIAMIIL---GLAIGIPLRAQRLODKG 1129
DB 2281 AGTASALASQLPQATMP---GKSMVALAGSSYOGNGLAIGV-----SRISDNG 2327

RESULT 15
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-16
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match          3.0%; Score 189.5; DB 4; Length 2736;
Best Local Similarity 20.8%; Pred. No. 0.00032;

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 11:46:12 : Search time 800 Seconds
(without alignments)
10343.567 Million cell updates/sec

Title: US-09-901-572A-2

Perfect score: 3189
Sequence: 1 atgaatatcttcaaaaact.....caacgcctcccaagaataa 3189

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published_Applications_MA:*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCOT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCOTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3189	100.0	3189	12	US-09-901-572A-2
2	3189	100.0	3189	14	US-10-131-591A-2
3	3143.4	95.6	8354	13	US-10-125-818-1
4	3057.8	95.9	3129	14	US-10-131-591A-79
5	124.6	3.9	580073	12	US-10-205-230-1
6	62.6	2.0	4985	12	US-10-056-405-10
7	62.6	2.0	4985	14	US-10-094-240-10
8	60.8	1.9	8771	12	US-10-311-455-1798
9	60.2	1.9	3489	12	US-10-294-804-1
10	58.4	1.8	640681	10	US-09-790-988-1
11	53.6	1.7	3673778	12	US-10-312-841-1
12	52.4	1.6	3057	12	US-10-349-680-148
13	51.6	1.6	2451	12	US-10-091-007-55
14	50.8	1.6	2017	14	US-10-155-533-3
15	50.8	1.6	5314	14	US-10-155-533-1
16	50.2	1.6	2010	9	US-09-878-756-3

c 17	49.4	1.5	1887	12	US-10-349-680-133	Sequence 133, App
c 18	47.8	1.5	3931	14	US-10-006-780-1	Sequence 1, Appli
c 19	47.6	1.5	1220	12	US-10-169-710-46	Sequence 46, Appli
c 20	47	1.5	575	11	US-09-918-995-6127	Sequence 6127, Ap
c 21	47	1.5	712	9	US-09-822-849A-186	Sequence 186, App
c 22	47	1.5	722	9	US-09-923-304-1	Sequence 1, Appli
c 23	47	1.5	722	10	US-09-880-107-1764	Sequence 1764, Ap
c 24	47	1.5	755	13	US-10-027-632-128109	Sequence 128109,
c 25	47	1.5	888	9	US-09-925-302-116	Sequence 116, App
c 26	47	1.5	2405	11	US-09-298-523B-70	Sequence 70, Appli
c 27	46.6	1.5	6351	12	US-10-311-455-1717	Sequence 1717, Ap
c 28	46.6	1.5	6351	12	US-10-240-485-139	Sequence 139, App
c 29	46.6	1.5	8634	12	US-10-311-455-1029	Sequence 1029, Ap
c 30	46.6	1.5	3673778	12	US-10-312-841-2	Sequence 2, Appli
c 31	46.4	1.5	11155	12	US-10-311-455-578	Sequence 578, App
c 32	46.2	1.4	5641	12	US-10-311-455-1369	Sequence 1369, Ap
c 33	46	1.4	860	9	US-09-770-445-596	Sequence 596, App
c 34	46	1.4	5304	12	US-10-240-453-229	Sequence 229, App
c 35	45.8	1.4	5285	12	US-10-240-453-1753	Sequence 1753, Ap
c 36	45.8	1.4	5285	12	US-10-240-453-177	Sequence 177, App
c 37	45.8	1.4	5285	14	US-10-239-676-157	Sequence 157, App
c 38	45.8	1.4	6171	12	US-10-311-455-761	Sequence 761, App
c 39	45.8	1.4	6644	12	US-10-311-455-488	Sequence 488, App
c 40	45.8	1.4	6641	12	US-10-311-455-288	Sequence 288, App
c 41	45.8	1.4	6641	12	US-10-240-453-36	Sequence 36, Appli
c 42	45.8	1.4	15732	12	US-10-240-453-108	Sequence 108, App
c 43	45.8	1.4	15732	14	US-10-239-676-96	Sequence 96, Appli
c 44	45.6	1.4	6151	12	US-10-311-455-1584	Sequence 1584, Ap
c 45	45.6	1.4	6151	12	US-10-240-453-150	Sequence 150, App

ALIGNMENTS

```
RESULT 1
US-09-901-572A-2
: Sequence 2, Application US/09901572A
: Publication No. US20030165534A1
: GENERAL INFORMATION:
: APPLICANT: Nippon Zeon Co., Ltd.,
: TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof
: FILE REFERENCE: J209
: CURRENT APPLICATION NUMBER: US/09/901,572A
: CURRENT FILING DATE: 2003-03-11
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: Patentln Ver. 2.1
: SEQ ID NO 2
: LENGTH: 3189
: TYPE: DNA
: ORGANISM: Mycoplasma gallisepticum
: FEATURE:
: OTHER INFORMATION: mgc3 gene.
US-09-901-572A-2
```

QY	1	ATGAAATATTTCTAATAAACTTAATAAGTTATACATTGATGAGTGATTAAGCTGATATTGGA	60
DB	1	ATGAATATTTCTAATAAACTTAATAAGTTATACATTGATGAGTGATTAAGCTGATATTGGA	60
QY	61	GCTCTTGTTCTGCAAGCTTTGGCTTTAAGCAATCGATTAAGAGTAAACGATTAACAGGCA	120
DB	61	GCTCTTGTTCTGCAAGCTTTGGCTTTAAGCAATCGATTAAGAGTAAACGATTAACAGGCA	120
QY	121	TTAGTTAATCAAGCAAGCAAGCTAGATGCTAATTCGTTAGACTTCAGCTTGGACAA	180
DB	121	TTAGTTAATCAAGCAAGCAAGCTAGATGCTAATTCGTTAGACTTCAGCTTGGACAA	180
QY	181	AATGTTGCTTGTCTTCAATACAGTTCTTAGAGATGTTGATGATTAACCTTTAACAGCAGCT	240
DB	181	AATGTTGCTTGTCTTCAATACAGTTCTTAGAGATGTTGATGATTAACCTTTAACAGCAGCT	240

QY	241	AATGGAAACATTATCAAAATTAGACTGTTTACTAAACCATTTATATGGTTAGATCAAGT	300
Db	241	AATGGAAACATTATCAAAATTAGACTGTTTACTAAACCATTTATATGGTTAGATCAAGT	300
QY	301	GATGATTGGTGGTGATACAAAGTAAACAAATAGTTGAGATTACACAACTAGCAGAAAT	360
Db	301	GATGATTGGTGGTGATACCAAGTAAACAAATAGTTGAGATTACCAACACTAGCAGAAAT	360
QY	361	AGATTTGATCAAGACAAACAGAGCATTTATGCTCTGTGGTTAAATGATGAAGCTAAC	420
Db	361	AGATTTGATCAAGACAAACAGAGCATTTATGCTCTGTGGTTAAATGATGAAGCTAAC	420
QY	421	GTTCAATTAAAGAAATTAATACTAAGTCAAAATAGAAATGGTAAATGGAACCAACATTC	480
Db	421	GTTCAATTAAAGAAATTAATACTAAGTCAAAATAGAAATGGTAAATGGAACCAACATTC	480
QY	481	AAGTTTGAATTTGGTGGTGTGATTAATCCAGCTCAGCTGATTAATGATTTACGATGATGG	540
Db	481	AAGTTTGAATTTGGTGGTGTGATTAATCCAGCTCAGCTGATTAATGATTTACGATGATGG	540
QY	541	ACTAAATTTAATTTTACAAACCAAACTCAAGTGAAATTTGTTATGACTTCATTTTAAT	600
Db	541	ACTAAATTTAATTTTACAAACCAAACTCAAGTGAAATTTGTTATGACTTCATTTTAAT	600
QY	601	GCGCAATCTTAAGATTAAACCCGATTTGGATTAACCTTATACCTTAATCAAGAAAG	660
Db	601	GCGCAATCTTAAGATTAAACCCGATTTGGATTAACCTTATACCTTAATCAAGAAAG	660
QY	661	ATCTTACCAATGACGTCACACTGCAGTTGTCCTTGGCCAGTAGTACGTAGTGTGA	720
Db	661	ATCTTACCAATGACGTCACACTGCAGTTGTCCTTGGCCAGTAGTACGTAGTGTGA	720
QY	721	ACAAATGGTGATGATGGGATGTTGATGTGGGAATGGTCAAAATTAATCAATACAGATCT	780
Db	721	ACAAATGGTGATGATGGGATGTTGATGTGGGAATGGTCAAAATTAATCAATACAGATCT	780
QY	781	ATTGCTCAAACTAAAAACCATCTAGTATATCAAAATCTTTCAACTTTTAATCAGAGAGA	840
Db	781	ATTGCTCAAACTAAAAACCATCTAGTATATCAAAATCTTTCAACTTTTAATCAGAGAGA	840
QY	841	ATGCCGTGGCCAAACAAATAGATAGATCTCAATTTGATGTCAAGCATAGAAATTAAGA	900
Db	841	ATGCCGTGGCCAAACAAATAGATAGATCTCAATTTGATGTCAAGCATAGAAATTAAGA	900
QY	901	TCTTTTCAATTAGATGAATAATTTGTTTATCCAAATAGGACTGGTTCTGAAGAAATTA	960
Db	901	TCTTTTCAATTAGATGAATAATTTGTTTATCCAAATAGGACTGGTTCTGAAGAAATTA	960
QY	961	AATATTAACAAGATTAGCTACTGGAAGTTTGGCAAGCAACGAAGAATTTGATTTTGAC	1020
Db	961	AATATTAACAAGATTAGCTACTGGAAGTTTGGCAAGCAACGAAGAATTTGATTTTGAC	1020
QY	1021	ATACCCGGACATCCACAGATTCCTTAAAGAAAGATTCAGTTAAGATTTTCAAGACTA	1080
Db	1021	ATACCCGGACATCCACAGATTCCTTAAAGAAAGATTCAGTTAAGATTTTCAAGACTA	1080
QY	1081	TACTTAAACTCAAGTTAATCTTTATCAATTCATGGTGTAGTATTTATTTTTGGTACC	1140
Db	1081	TACTTAAACTCAAGTTAATCTTTATCAATTCATGGTGTAGTATTTATTTTTGGTACC	1140
QY	1141	TCGCAATTCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1200
Db	1141	TCGCAATTCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1200
QY	1201	TTGCAATCAAGTTAAAAACAATGATTTGAAGCTTCACACATGATTAACGATCAACA	1260
Db	1201	TTGCAATCAAGTTAAAAACAATGATTTGAAGCTTCACACATGATTAACGATCAACA	1260
QY	1261	AAGGAAACAAGCAACAGCTGATTAATGATGATGATGATGATGATGATGATGATGATG	1320
Db	1261	AAGGAAACAAGCAACAGCTGATTAATGATGATGATGATGATGATGATGATGATGATG	1320

QY	1321	ACTACTAACACCTTCTCAACACGTTTCTAATCTCTACTTAAATACCTATCGTACTGTTGGG	1380
Db	1321	ACTACTAACACCTTCTCAACACGTTTCTAATCTCTACTTAAATACCTATCGTACTGTTGGG	1380
QY	1381	ATTGATGTGAACCACTCTGCAACAAACAAATAGTGAACCTAATTGGGCGAGTCTCTAAC	1440
Db	1381	ATTGATGTGAACCACTCTGCAACAAACAAATAGTGAACCTAATTGGGCGAGTCTCTAAC	1440
QY	1441	GTTATTGAAGCAAGATATATGCTGAATPACAGATTAGGTATTCAAAATGGAATTTCCATA	1500
Db	1441	GTTATTGAAGCAAGATATATGCTGAATPACAGATTAGGTATTCAAAATGGAATTTCCATA	1500
QY	1501	ACTAATGAGGAACCTTTATCCGGAACCAATGTTGGTGGTGGTGGTTTACTTCCAAACGT	1560
Db	1501	ACTAATGAGGAACCTTTATCCGGAACCAATGTTGGTGGTGGTGGTGGTTTACTTCCAAACGT	1560
QY	1561	TCAAAGATAGTTTTAAGAGCTTTATAGGCTGATCAGCTCCAACTGGAAACTTCCAA	1620
Db	1561	TCAAAGATAGTTTTAAGAGCTTTATAGGCTGATCAGCTCCAACTGGAAACTTCCAA	1620
QY	1621	CCTTTCTTATACGTATTTGGTATTTAGGATPACCAACCAACTAGAACGGAACCTTTGCG	1680
Db	1621	CCTTTCTTATACGTATTTGGTATTTAGGATPACCAACCAACTAGAACGGAACCTTTGCG	1680
QY	1681	TACGGAACATATPACGTTTTAACAACAACGCTTACGACGATTTAGATTCTCCAAAGATA	1740
Db	1681	TACGGAACATATPACGTTTTAACAACAACGCTTACGACGATTTAGATTCTCCAAAGATA	1740
QY	1741	GGTACTGAACCAATCATTTTGAAGAACTTGTATTAACATACCCTGTTATGGGTGGATAT	1800
Db	1741	GGTACTGAACCAATCATTTTGAAGAACTTGTATTAACATACCCTGTTATGGGTGGATAT	1800
QY	1801	CTPACTGAAGAAGGCTCTAGAAGTTCTCTAATPACTCCATATATATAGAGCACAGGTGAC	1860
Db	1801	CTPACTGAAGAAGGCTCTAGAAGTTCTCTAATPACTCCATATATATAGAGCACAGGTGAC	1860
QY	1861	ACACCGAAAAAGCCGAAGCATCTTCCATCGGCTATTCGATTAATACCTATAGTACATTT	1920
Db	1861	ACACCGAAAAAGCCGAAGCATCTTCCATCGGCTATTCGATTAATACCTATAGTACATTT	1920
QY	1921	CATTCAGTTTATAGATTTGATGGAATTAGAAATAACTTAAATGTTGGGTTTAAAGCATCA	1980
Db	1921	CATTCAGTTTATAGATTTGATGGAATTAGAAATAACTTAAATGTTGGGTTTAAAGCATCA	1980
QY	1981	AGCTTCTTAACCTCAAAATAGACCAAAATCCAAACGGTCTAGAAATGATTTCTGCACACAAA	2040
Db	1981	AGCTTCTTAACCTCAAAATAGACCAAAATCCAAACGGTCTAGAAATGATTTCTGCACACAAA	2040
QY	2041	TACTTAAGATCACAAATGGAATTAGGTAGAGCAACTCTGGAATTACCAACCAACCAACCTTC	2100
Db	2041	TACTTAAGATCACAAATGGAATTAGGTAGAGCAACTCTGGAATTACCAACCAACCAACCTTC	2100
QY	2101	GGAACAACCTCAACAGTTATTTAGATACCTGCTGGATGATGTTCTCATCAATTAAGAAAT	2160
Db	2101	GGAACAACCTCAACAGTTATTTAGATACCTGCTGGATGATGTTCTCATCAATTAAGAAAT	2160
QY	2161	ATTGAAACAACCTTCCCTCGGTACCAATTTAGTACTTCTTATTCACAAATGAAAAATAT	2220
Db	2161	ATTGAAACAACCTTCCCTCGGTACCAATTTAGTACTTCTTATTCACAAATGAAAAATAT	2220
QY	2221	AAATCTAGTGTTTATACATTAAGATTAGCTGACTCAAGTAAACCTGATGCTCAAGCTCA	2280
Db	2221	AAATCTAGTGTTTATACATTAAGATTAGCTGACTCAAGTAAACCTGATGCTCAAGCTCA	2280
QY	2281	TTTCACTCCACAAAGTTTATTTAGAGCTTAATGAAGAAATTTGGTGAATCTTAACTTATTTAGAC	2340
Db	2281	TTTCACTCCACAAAGTTTATTTAGAGCTTAATGAAGAAATTTGGTGAATCTTAACTTATTTAGAC	2340
QY	2341	AATTCATCTCTPACAGTAAATGCTGCTGGTAAATGTTGCATTTGTTCTCATCAAAACCTGCT	2400
Db	2341	AATTCATCTCTPACAGTAAATGCTGCTGGTAAATGTTGCATTTGTTCTCATCAAAACCTGCT	2400
QY	2401	TCTCTGTGATCATATAGTCTGTGAATAATCAATTTAATCAGAACCTTATCTGATATTCGTTTT	2460
Db	2401	TCTCTGTGATCATATAGTCTGTGAATAATCAATTTAATCAGAACCTTATCTGATATTCGTTTT	2460

Db	2401	TCCTCGGATCATATACTGCTGTAATATACATTTAATCAGAACTTATCTGATTTGCTTTT	2460
Qy	2461	GAAGGTTCTGGTGGTAAAGTATACATCTGATTTCTGGGGAAACATCCAAATTCAAACCGGAT	2520
Db	2461	GAAGGTTCTGGTGGTAAAGTATACATCTGATTTCTGGGGAAACAAATTCAAATCCGAT	2520
Qy	2521	GAGTACTTAATTCAAAATGGGTTCTACATGATCAAGTGGGCTAGAAACCTTGTTTACAAACCA	2580
Db	2521	GAGTACTTAATTCAAAATGGGTTCTACATGATCAAGTGGGCTAGAAACCTTGTTTACAAACCA	2580
Qy	2581	AGCTTCTTAAACAGTTTAGTTGACTTCCTCGTCTAATGCTGGTACTAATCACCCTGTA	2640
Db	2581	AGCTTCTTAAACAGTTTAGTTGACTTCCTCGTCTAATGCTGGTACTAATCACCCTGTA	2640
Qy	2641	GTGGTTGATCTCGTAAATTTTAAACAAACCAAACTTACTCTTAAAGTTCAATGCCA	2700
Db	2641	GTGGTTGATCTCGTAAATTTTAAACAAACCAAACTTACTCTTAAAGTTCAATGCCA	2700
Qy	2701	TACTTAGATGTAAGTATTTATGATGCTAAATTTAAGAACAATATTTAGTAAATCTCT	2760
Db	2701	TACTTAGATGTAAGTATTTATGATGCTAAATTTAAGAACAATATTTAGTAAATCTCT	2760
Qy	2761	TATTAACAATTTGGCGCCTTACCTTCATTCGGTAGTGCCTTACAGCAATTTGGTAGTACATTA	2820
Db	2761	TATTAACAATTTGGCGCCTTACCTTCATTCGGTAGTGCCTTACAGCAATTTGGTAGTACATTA	2820
Qy	2821	GGTATTCCTGCAATTAATGATCATCTTAGAGATTAGCTATTCGGTATTTCTTTAAGAGCTCAA	2880
Db	2821	GGTATTCCTGCAATTAATGATCATCTTAGAGATTAGCTATTCGGTATTTCTTTAAGAGCTCAA	2880
Qy	2881	AGAAATTTACAGACAAAGGTTTCAAAACAATTCAAAAGTTGATACCTTGAGCT	2940
Db	2881	AGAAATTTACAGACAAAGGTTTCAAAACAATTCAAAAGTTGATACCTTGAGCT	2940
Qy	2941	GCTGTTGGTTCAGTTTCAAGAAAGATTATTAACCAACTGCTAACGTTAAGAAAAAAGCT	3000
Db	2941	GCTGTTGGTTCAGTTTCAAGAAAGATTATTAACCAACTGCTAACGTTAAGAAAAAAGCT	3000
Qy	3001	GCTGCTTTAGTGCTGTTAAATCTGTTGATTAAGAAACCTGCTGCTGCTAAACCTGCT	3060
Db	3001	GCTGCTTTAGTGCTGTTAAATCTGTTGATTAAGAAACCTGCTGCTGCTGCTAAACCTGCT	3060
Qy	3061	GCTCCAGCTTAACCATCTGCACCAAAAGCTAAGCTCAGCTAAACCAACTGGGCTTAA	3120
Db	3061	GCTCCAGCTTAACCATCTGCACCAAAAGCTAAGCTCAGCTAAACCAACTGGGCTTAA	3120
Qy	3121	TCTGTGTCGCTTACAAACCAACCACTGCTCTTAAGCAGCTGCTCCAAAACCAACGCTGCC	3180
Db	3121	TCTGTGTCGCTTACAAACCAACCACTGCTCTTAAGCAGCTGCTCCAAAACCAACGCTGCC	3180
Qy	3181	AAAGATATTA 3189	
Db	3181	AAAGATATTA 3189	
RESULT 2			
US-10-131-591A-2			
: Sequence 2, Application US/10131591A			
: Publication NO. US20030059799A1			
: GENERAL INFORMATION:			
: APPLICANT: Nippon Zeon Co., Ltd.,			
: TITLE OF INVENTION: Modified DNA molecules, Recombinants and uses thereof			
: FILE REFERENCE: J709			
: CURRENT APPLICATION NUMBER: US/10/131,591A			
: NUMBER OF SEQ ID NOS: 79			
: SOFTWARE: PatentIn Ver. 2.1			
: SEQ ID NO 2			
: LENGTH: 3189			
: TYPE: DNA			
: ORGANISM: Mycoplasma gallisepticum			
: FEATURE:			

OTHER INFORMATION: mgc3 gene									
US-10-131-591A-2									
Query Match	100.0%	Score 3189	DB 14	Length 3189					
Best Local Similarity	100.0%	Prod. No. 0							
Matches 3189	Conservative 0	Mismatches 0	Indels 0	Gaps 0					
Oy	1	ATGAATATTTCTCTAAAAA	CTTAAAGATTTATCATTTAGTGGATTTAGCTGTAATTTGA	60					
Db	1	ATGAATATTTCTTAAAAA	CTTAAAGATTTATCATTTAGTGGATTTAGCTGTAATTTGA	60					
Oy	61	GCTCTGTGTTCTGCAAGCTTT	GCGTTTAAGCAATCAGATAAGACTAACGATTAACGACAA	120					
Db	61	GCTCTGTGTTCTGCAAGCTTT	GCGTTTAAGCAATCAGATAAGACTTAAGCAATTAACGACAA	120					
Oy	121	TTAGTTAATCAAGCAAGACCT	TGATGCTTAATTTCTGTTAGACTTTGCAGCTTGTGACAA	180					
Db	121	TTAGTTAATCAAGCAAGACCT	TGATGCTTAATTTCTGTTAGACTTTGCAGCTTGTGACAA	180					
Oy	181	AATGGTTCGTTTCAATACAGT	CTTGAAGATGTTGATGATTAACCTTATTAACAGACCT	240					
Db	181	AATGGTTCGTTTCAATACAGT	CTTGAAGATGTTGATGATTAACCTTATTAACAGACCT	240					
Oy	241	AATGGAACAATTTATCAAA	TAGATGTTTACTTAACCACTTAATATGTTTAGATCTAAGT	300					
Db	241	AATGGAACAATTTATCAAA	TAGATGTTTACTTAACCACTTAATATGTTTAGATCTAAGT	300					
Oy	301	GATGATGTGTTGATACAAAG	TAAACCAATATGTTTCAGATTCACACACTGACGAAT	360					
Db	301	GATGATGTGTTGATACAAAG	TAAACCAATATGTTTCAGATTCACACACTGACGAAT	360					
Oy	361	AGATTTGTATCAAGCAAG	ACATATTTATGCTCTGTTGGTTAATGATGAAGCTAAC	420					
Db	361	AGATTTGTATCAAGCAAG	ACATATTTATGCTCTGTTGGTTAATGATGAAGCTAAC	420					
Oy	421	GTCATTTTAAAAAGATTA	TATACATACCAATGGAATTTGGTATATAGAAACAACAATTTCT	480					
Db	421	GTCATTTTAAAAAGATTA	TATACATACCAATGGAATTTGGTATATAGAAACAACAATTTCT	480					
Oy	481	AAGTTTGTAAATTTGGT	GTGATTAATTCACAGCTCACGTAATTAGATTACTGATGATGG	540					
Db	481	AAGTTTGTAAATTTGGT	GTGATTAATTCACAGCTCACGTAATTAGATTACTGATGATGG	540					
Oy	541	ACTAAATTTAATTTTACA	AACCACTCAGAGTGAATTTGTTAATGACTTATTTTAGAT	600					
Db	541	ACTAAATTTAATTTTACA	AACCACTCAGAGTGAATTTGTTAATGACTTATTTTAGAT	600					
Oy	601	GGGCAATCTTACCTAAAG	ATTTTACCCAGATTTGGTTTAACTTATFACATTTCAAGAAAG	660					
Db	601	GGGCAATCTTACCTAAAG	ATTTTACCCAGATTTGGTTTAACTTATFACATTTCAAGAAAG	660					
Oy	661	ATCTTTACCAATTTGAG	CGTCACGTCAGATTTGCTCTTGGCCAGTAGAGTAGATTAGTGA	720					
Db	661	ATCTTTACCAATTTGAG	CGTCACGTCAGATTTGCTCTTGGCCAGTAGAGTAGATTAGTGA	720					
Oy	721	ACCAATGCTGATGATGG	ATGTTGATTTGGGGAATGGTCAAAATTAACATTAACAGATCT	780					
Db	721	ACCAATGCTGATGATGG	ATGTTGATTTGGGGAATGGTCAAAATTAACATTAACAGATCT	780					
Oy	781	ATTCGCTCAAACTAAAA	CACACTACGATTAATCAAAATCCTTCAACTTTAATTCAGSAGCA	840					
Db	781	ATTCGCTCAAACTAAAA	CACACTACGATTAATCAAAATCCTTCAACTTTAATTCAGSAGCA	840					
Oy	841	ATGCGTGTGCAAAAC	AATATAGATTTCTCAATTTGAATGTCAAGCATATGAATTAACA	900					
Db	841	ATGCGTGTGCAAAAC	AATATAGATTTCTCAATTTGAATGTCAAGCATATGAATTAACA	900					
Oy	901	TCTTTCCAAATTAGATGA	AAAAATTTGTTTATCCAGAAATGACCTGGTTCTGAAAGAGATTA	960					
Db	901	TCTTTCCAAATTAGATGA	AAAAATTTGTTTATCCAGAAATGACCTGGTTCTGAAAGAGATTA	960					
Oy	961	AATATTAACAAGATTAG	CTACTGGAATTTGCCAAGCAAGAAATATTTGATTTCTTGAC	1020					
Db	961	AATATTAACAAGATTAG	CTACTGGAATTTGCCAAGCAAGAAATATTTGATTTCTTGAC	1020					

Db	961	AATATTACAAAGATTAGTACTGGAAGTTGCCAACAAAGAAAGATTAATGATCTTCTAC	1020
Qy	1021	ATACCGGGGACTCCAAAGTTACTTTAAAGAAAGTTCAGTTAACTATTTTTCAGAGCTA	1080
Db	1021	ATACCGGGGACTCCAAAGTTACTTTAAAGAAAGTTCAGTTAACTATTTTTCAGAGCTA	1080
Qy	1081	TACTTAACCTCAGTTAATCTTTATCATTTCATTTGTTGATAGTATTTATTTTGGTACC	1140
Db	1081	TACTTAACCTCAGTTAATCTTTATCATTTCATTTGTTGATAGTATTTATTTTGGTACC	1140
Qy	1141	TCGGAATTCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1200
Db	1141	TCGGAATTCACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1200
Qy	1201	TTGGAATTCAGTTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1260
Db	1201	TTGGAATTCAGTTAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1260
Qy	1261	AACGGAACAAACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1320
Db	1261	AACGGAACAAACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1320
Qy	1321	ACTACTACACTCTCTCAACAGTTCTTAATCTTAATCTTAATCTTAATCTTAATCTTAATCT	1380
Db	1321	ACTACTACACTCTCTCAACAGTTCTTAATCTTAATCTTAATCTTAATCTTAATCTTAATCT	1380
Qy	1381	ATTGATGATTAACCAACTCTCTCAACAGTTCTTAATCTTAATCTTAATCTTAATCTTAATCT	1440
Db	1381	ATTGATGATTAACCAACTCTCTCAACAGTTCTTAATCTTAATCTTAATCTTAATCTTAATCT	1440
Qy	1441	GTTATTTGAACCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
Db	1441	GTTATTTGAACCAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1500
Qy	1501	ACTAATCAGAGAACTTTATCGAAACAACTATGATGATGATGATGATGATGATGATGATGAT	1560
Db	1501	ACTAATCAGAGAACTTTATCGAAACAACTATGATGATGATGATGATGATGATGATGATGAT	1560
Qy	1561	TCGAAGTATGTTTAAAGCTTTCTTAATGATGATGATGATGATGATGATGATGATGATGAT	1620
Db	1561	TCGAAGTATGTTTAAAGCTTTCTTAATGATGATGATGATGATGATGATGATGATGATGAT	1620
Qy	1621	CCCTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1680
Db	1621	CCCTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1680
Qy	1681	TACGGAACATTAAGCTTTTAAACACAGCTTACAGCTTATGATGATGATGATGATGATGATGAT	1740
Db	1681	TACGGAACATTAAGCTTTTAAACACAGCTTACAGCTTATGATGATGATGATGATGATGATGAT	1740
Qy	1741	GATACGTAACCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1800
Db	1741	GATACGTAACCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1800
Qy	1801	CTAATGTAAGAGTGTCTAGAAAGTTCTTAATGATGATGATGATGATGATGATGATGATGAT	1860
Db	1801	CTAATGTAAGAGTGTCTAGAAAGTTCTTAATGATGATGATGATGATGATGATGATGATGAT	1860
Qy	1861	ACACGAAAGCCGGAAGCTTCCAACTGCTATTCGATGATGATGATGATGATGATGATGATGAT	1920
Db	1861	ACACGAAAGCCGGAAGCTTCCAACTGCTATTCGATGATGATGATGATGATGATGATGATGAT	1920
Qy	1921	CAATCAGTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1980
Db	1921	CAATCAGTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1980
Qy	1981	AGCTTCTTAACCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2040
Db	1981	AGCTTCTTAACCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2040
Qy	2041	TACTTAAGTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2100
Db	2041	TACTTAAGTATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2100

Qy	2101	GGAACACATCACCAGTTATTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT	2160
Db	2101	GGAACACATCACCAGTTATTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGAT	2160
Qy	2161	ATTAGAACCAATCTCCCTGGTAAACAGTTATGATGATGATGATGATGATGATGATGATGAT	2220
Db	2161	ATTAGAACCAATCTCCCTGGTAAACAGTTATGATGATGATGATGATGATGATGATGATGAT	2220
Qy	2221	AAATCTAGTGTGTTTACATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT	2280
Db	2221	AAATCTAGTGTGTTTACATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT	2280
Qy	2281	TTTCACTCAACCAAGTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2340
Db	2281	TTTCACTCAACCAAGTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2340
Qy	2341	AATTCATCTTAACAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2400
Db	2341	AATTCATCTTAACAGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2400
Qy	2401	TCTCTGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2460
Db	2401	TCTCTGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2460
Qy	2461	GAAGGTTCTGCTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2520
Db	2461	GAAGGTTCTGCTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2520
Qy	2521	GAGTCTTAATTCAAAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT	2580
Db	2521	GAGTCTTAATTCAAAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT	2580
Qy	2581	AGCTTCTTAACCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2640
Db	2581	AGCTTCTTAACCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2640
Qy	2641	GTTGTTGATCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2700
Db	2641	GTTGTTGATCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2700
Qy	2701	TACTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2760
Db	2701	TACTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2760
Qy	2761	TATTAACACTTGGGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT	2820
Db	2761	TATTAACACTTGGGCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT	2820
Qy	2821	GATATCTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2880
Db	2821	GATATCTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	2880
Qy	2881	AGAAATTAACAAAGAGGTTCAAAACAACTTCAAAACAACTTCAAAACAACTTCAAAACAACT	2940
Db	2881	AGAAATTAACAAAGAGGTTCAAAACAACTTCAAAACAACTTCAAAACAACTTCAAAACAACT	2940
Qy	2941	GCTGTTGTTCACTTTTCAAGAAAGATTAATACCAAACTGCTAAGTTAAGAAAAAAGCT	3000
Db	2941	GCTGTTGTTCACTTTTCAAGAAAGATTAATACCAAACTGCTAAGTTAAGAAAAAAGCT	3000
Qy	3001	GCTGTTTAAAGTGTGTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3060
Db	3001	GCTGTTTAAAGTGTGTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3060
Qy	3061	GCTTCAAGCTTAACCTGCTGCAAAAGTACCTGCAAAAGTACCTGCAAAAGTACCTGCTGCA	3120
Db	3061	GCTTCAAGCTTAACCTGCTGCAAAAGTACCTGCAAAAGTACCTGCAAAAGTACCTGCTGCA	3120
Qy	3121	TCTGTTGCTGCTCAAAACCAATGCTGCTCAAAACCAATGCTGCTCAAAACCAATGCTGCT	3180
Db	3121	TCTGTTGCTGCTCAAAACCAATGCTGCTCAAAACCAATGCTGCTCAAAACCAATGCTGCT	3180

QY	3181	AAAGATTA	3189
Db	3181	AAAGATTA	3189

```

RESULT 3
US-10-125-818-1
: Sequence 1, Application US/10125818
: Publication NO. US20020187162A1
: GENERAL INFORMATION:
: APPLICANT: University of Connecticut
: APPLICANT: Geary, Steven J.
: APPLICANT: Sekelluck, Margaret
: APPLICANT: Marcus, Phillip J.
: APPLICANT: Silbart, Lawrence
: TITLE OF INVENTION: Use of a Live Attenuated Mycoplasma gallisepticum Strain as a Vaccine
: TITLE OF INVENTION: Vector for the Protection of Chickens and Turkeys from Respiract
: FILE REFERENCE: 883993.0074
: CURRENT APPLICATION NUMBER: US/10/125.818
: CURRENT FILING DATE: 2002-04-19
: PRIOR APPLICATION NUMBER: 60/285,569
: PRIOR FILING DATE: 2001-04-21
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1
: LENGTH: 8354
: TYPE: DNA
: ORGANISM: Mycoplasma gallisepticum
: US-10-125-818-1

```

Query Match	98.6%;	Score 3143.4;	DB 13;	Length 8354;
Best Local Similarity	99.1%;	Pred. No. 0;		
Matches 3159; Conservative	2;	Mismatches	28;	Indels 0;

QY	1	ATGAATATTTCTAAAAAACTTAAAAAGTTATACATTTGATAGGTGGATTAAGCTGTATTTGGA	60
Db	3778	ATGAATATTTCTAAAAAACTTAAAAAGTTATACATTTGATAGGTGGATTAAGCTGTATTTGGA	3837
OY	61	GCTCTTGGTTCGAAGCTTTGGCTTTTAAAGCAATCAGATTAGAGTAAGCAAGTAACCGAA	120
Db	3838	GCTCTTGGTTCGAAGCTTTGGCTTTTAAAGCAATCAGATTAGAGTAAGCAAGTAACCGAA	3897
OY	121	TTAGTTAATCAGACAGAACGCTAGATGCTAATTCCTGTAGACTTCGACGTCCTTGACAA	180
Db	3898	TTAGTTAATCAGACAGAACGCTAGATGCTAATTCCTGTAGACTTCGACGTCCTTGACAA	3955
OY	181	AATGCTTCGTTGTTCAATACACGTTCTTAGAGATGTTGATGATAACTTTATATACAGCAGCT	240
Db	3958	AATGCTTCGTTGTTCAATACACGTTCTTAGAGATGTTGATGATAACTTTATATACAGCAGCT	4017
OY	241	AATGGAACAATATTAACAATTGATAGATTTTCTAACAATATATAGTTTGGTTTGGATCTAAGT	300
Db	4018	AATGGAACAATATTAACAATTGATAGATTTTCTAACAATATATAGTTTGGTTTGGATCTAAGT	4077
OY	301	GATGATTTGGTGATACAAAGTAAACAAATAGTTTCAGATTCACACAATCAGACGAAT	360
Db	4078	GATGATTTGGTGATACAAAGTAAACAAATAGTTTCAGATTCACACAATCAGACGAAT	4137
OY	361	AGATTTGATCAACAGCAACAAGACATATTATGCTCTGTTGGTTAATGATGAAGCTAAC	420
Db	4138	AGATTTGATCAACAGCAACAAGACATATTATGCTCTGTTGGTTAATGATGAAGCTAAC	4197
OY	421	GTCATATTAAAGAAATTAATACAGAACCAATAGAAATGGTAATAGAAACAAATTCCT	480
Db	4198	GTCATATTAAAGAAATTAATACAGAACCAATAGAAATGGTAATAGAAACAAATTCCT	4257
OY	481	AAGTTTGGTAATGGTGTTGATGAATTCACACTCAGCAATATAGATTACTGATGATGGG	540
Db	4258	AAGTTTGGTAATGGTGTTGATGAATTCACACTCAGCAATATAGATTACTGATGATGGG	4317
OY	541	ACTAAATTTAATTTTACAAACCAACCTCAGGTGAATTTGTTAATGACTTCATTTTACAT	600
Db	4318	ACTAAATTTAATTTTACAAACCAACCTCAGGTGAATTTGTTAATGACTTCATTTTACAT	4377

OY	601	GCGCCATCTTACCTAAAAGATTTACACCCAGATTGGTATTAACATTATACATTCCAAGAAAG	660
Db	4378	GGCGCAATCTTACCTAAAAGATTTACACCCAGATTGGTATTAACATTATACATTCCAAGAAAG	4437
OY	661	ATCTTACCAATGACGTCACACACACGAGTTTCTCTGGCCAGTAGAGGATGATGAGGA	720
Db	4438	ATCTTACCAATGACGTCACACACGAGTTTCTCTGGCCAGTAGAGGATGATGAGGGA	4497
OY	721	ACAAATGCTGATGATGGGATGTTGATTTGGGAATGCTCAATTAACATATACAGATCT	780
Db	4498	ACAAATGCTGATGATGGGATGTTGATTTGGGAATGCTCAATTAACATATACAGATCT	4557
OY	781	ATGCGTCAAAATTAACCACTACGATGATTAATCAAAATCCTTCACATTTTAAATTCGAGACA	840
Db	4558	ATGCGTCAAAATTAACCACTACGATGATTAATCAAAATCCTTCACATTTTAAATTCGAGACA	4617
OY	841	ATGCGTGGTGCMAACAATAGATAGATTTCTCAATTGGAATGTCAAGCATAGATTTAAACA	900
Db	4618	ATGCGTGGTGCMAACAATAGATAGATTTCTCAATTGGAATGTCAAGCATAGATTTAAACA	4677
OY	901	TCTTTCCAAATAGATGAAAAATTTGGTTATCCACAATGAGACTGGTCTCGAAGAAATPAA	960
Db	4678	TCTTTCCAAATAGATGAAAAATTTGGTTATCCACAATGAGACTGGTCTCGAAGAAATPAA	4737
OY	961	AATATTACAGATTAGTACTGAGAAAGTTTGGCAAGCAAGAAAGATTTGGATTCCTTGAC	1020
Db	4738	AATATTACAGATTAGTACTGAGAAAGTTTGGCAAGCAAGAAAGATTTGGATTCCTTGAC	4797
OY	1021	ATACCCGGGACCTCCACAAGTTACTTTAAAGAAAGTTCAGTTAACGTRATTTTCAAGACTA	1080
Db	4798	ATACCCAGGAGCTCCACAAGTTCCTTTAAAGAAAGTTCAGTTAACGTRATTTTCAAGACTA	4857
OY	1081	TACTTAAACCTCAGTTAATCTTATATCAATTCATTCATGCTGTGATAGTATTATTTTGGTACC	1140
Db	4858	TACTTAAACCTCAGTTAATCTTATATCAATTCATTCATGCTGTGATAGTATTATTTTGGTACT	4917
OY	1141	TCGTGAATTACATCTTATATGCTACTATTATTCCTCAACTAGATTAATCTGATCTAACCGCT	1200
Db	4918	TCGTGAATTACATCTTATATGCTACTATTATTCCTCAACTAGATTAATCTGATCTAACCGCT	4977
OY	1201	TTGATATCAAGTTAAAAAGATGATATGATTAACCTTCAAGCACTGATGATTAACGGTACAAACA	1260
Db	4978	TTGATATCAAGTTAAAAAGATGATATGATTAACCTTCAAGCACTGATGATTAACGGTACAAACA	5037
OY	1261	AACGGAAACACAGACAAACAGCTGATACATCTAGTGGTTCACAGGTCGTCGAACAGAAAT	1320
Db	5038	AACGGAAACACAGACAAACAGCTGATACATCTAGTGGTTCACAGGTCGTCGAACAGAAAT	5097
OY	1321	ACTACTAACACTTCTCAACAAGTTTCTATCTCTACTTTAAATFACTTATCGTAGTTTGGGA	1380
Db	5098	ACTACTAACACTTCTCAACAAGTTTCTATCTCTACTTTAAATFACTTATCGTAGTTTGGGA	5157
OY	1381	ATTATATAGTAACCAACTCTGCAAAACAAAAATAGATGAAACTAAATTTGGGCGAGATCTTAAC	1440
Db	5158	ATTATATAGTAACCAACTCTGCAAAACAAAAATAGATGAAACTAAATTTGGGCGAGATCTTAAT	5217
OY	1441	GTTATTGAGCAAGAAATATATGCTGATACAGATTAAGTATTTCAAAATGAATTCACAATA	1500
Db	5218	GTTATTGAGCAAGAAATATATGCTGATACAGATTAAGTATTTCAAAATGAATTCACAATA	5277
OY	1501	ACTAATGACGAAAACTTATTCGGAACCAATGGTGGTGTGGTTTACTTCAACAGGT	1560
Db	5278	ACTAATGACGAAAACTTATTCGGAACCAATGGTGGTGTGGTTTACTTCAACAGGT	5337
OY	1561	TCAAGAGTAGTTTTAAGAGCTCTTATTAACGGTATCAACGCTCCAACTGGAACCTTCCAA	1620
Db	5338	TCAAGAGTAGTTTTAAGAGCTCTTATTAACGGTATCAACGCTCCAACTGGAACCTTCCAA	5397
OY	1621	CCTTTCTTATACGATTTGGTTATTTAGGATACCAACAAACTAGAACAGAACTTTCGG	1680
Db	5398	CCTTTCTTATACGATTTGGTTATTTAGGATACCAACAAACTAGAACAGAACTTTCGGA	5457

Db	6538	TATAACAACCTTGGCTGCTTACCTTCATGATGATGCTTACAGCAATTTGGTACTACATTTA	6597
QY	2821	GGTATCTCTGCAATTAATGATCATCTTAGATTAGCTATCGTATTCCTTTAAGACTCA	2880
Db	6598	GGTATCTCTGCAATTAATGATCATCTTAGATTAGCTATCGTATTCCTTTAAGACTCA	6657
QY	2881	AGAAATATTCAGACAAAGGGTTCCAAACACATTCGAAAAAGTTGATACCTTGACTGCT	2940
Db	6658	AGAAATATTCAGACAAAGGGTTCCAAACACATTCGAAAAAGTTGATACCTTGACTGCT	6717
QY	2941	GCTGTGGTGAAGTTCCAGAGAGATTATTAACCAACTGCTAACGTTAGAAAAAACT	3000
Db	6718	GCTGTGGTGAAGTTCCAGAGAGATTATTAACCAACTGCTAACGTTAGAAAAAACT	6777
QY	3001	GCTGCTTAAAGTGGTGGTAAATCTGTGATAGAAACCTGCTGCTGCTTAAACCTGCT	3060
Db	6778	GCTGCTTAAAGTGGTGGTAAATCTGTGATAGAAACCTGCTGCTGCTTAAACCTGCT	6837
QY	3061	GCTCCAGCTAAACCATCTCTGCACCAAAAGCTAGCTACCGCTAAACCAACTGGGCTTAA	3120
Db	6838	GCTCCAGCTAAACCATCTCTGCACCAAAAGCTAGCTACCGCTAAACCAACTGGGCTTAA	6897
QY	3121	TCTGTTGGGCGCTACGAAAAACAAGTCTCTTAAGCCAGCTGCTCCAAAACCGCTGCC	3180
Db	6898	TCTGTTGGGCGCTACGAAAAACAAGTCTCTTAAGCCAGCTGCTCCAAAACCGCTGCC	6957
QY	3181	AAAGAAATAA 3189	
Db	6958	AAAGAAATAA 6966	

[illegible]


```

? RESULT 5
? US-10-205-220-1
? : Sequence 1, Application US/10205220
? : Publication No. US20030170663A1
? :
? : GENERAL INFORMATION:
? : APPLICANT: Frazer et al.
? : TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma genitalium Genome, Fragment
? : TITLE OF INVENTION: Thereof, and Uses Thereof
? : FILE REFERENCE: PBI93PID1
? : CURRENT APPLICATION NUMBER: US/10/205,220
? : CURRENT FILING DATE: 2002-07-26
? : PRIOR APPLICATION NUMBER: US 08/545,528
? : PRIOR FILING DATE: 1995-10-19
? : PRIOR APPLICATION NUMBER: US 08/488,018
? : PRIOR FILING DATE: 1995-06-07
? : PRIOR APPLICATION NUMBER: US 08/473,545
? : PRIOR FILING DATE: 1995-06-07
? : NUMBER OF SEQ ID NOS: 1
? : SOFTWARE: PatentIn version 3.1
? : SEQ ID NO 1
? : LENGTH: 580073
? :
? : TYPE: DNA
? :
? : ORGANISM: Mycoplasma genitalium

```

QY	2770	TTTGGGCGCCTTACCTTCATGGGTAGTGGCCCTACAGCAATTTGGTACTACATTAGATTTCTT	2829
Db	228703	TTTGCACACCTCCCTCCTACGGGTGATGCCCTGTATAGTAGGTTCTCAGTTGGGATCTTG	228762
QY	2830	GCAATTATGATTCATCTTAGGATAGCTATCGGTATTCCTTTAAGACTCAAAAGAAATT	2889
Db	228763	TTTATCTGTATAGCTTAGGACTTGGGATTTGGATCCCAATGTACAGGGTAAAGAAATC	228822
QY	2890	CAACAACAAGGTTCTCAAAACAACATTCACAAAAAGTTGATACCTTGACTGCTGTGGT	2949
Db	228823	CAGAATGCACTGTTTGTATTATGCTTTAAAGGTGATACACTCCACAACCTGCTGGT	228882
QY	2950	TCAGTTTACAAGANATTTATTACCCAAACCTGCAACGTTAAACAAAAAACCTGCGCTTTA	3009
Db	228883	AGTGTGTACAAAAAGATTATTATCCCAACCTGGTGTGTAAAAAAGCAACCTAGTGCATTTG	228941
QY	3010	GGTGCTGTAAATCTGGGTATAGAAACCTGCTGCTGCTGCTAAACCTGCTGCTCTCAC	3068
Db	228943	AAAGTGTGTAATCTAGTGTAAAAAACCTGCGCTTTTATTAAACCACTGTTTAAAC	229001

```

1  APPLICANT ZWIEBEL, LAURENCE J.
2  TITLE OF INVENTION: MOSQUITO OLFACATORY GENES, POLYPEPTIDES, AND METHODS OF
3  TITLE OF INVENTION:  USE THEREOF
4  FILE REFERENCE: n7841
5  CURRENT APPLICATION NUMBER: US/10/056,405
6  CURRENT FILING DATE: 2002-01-24
7  PRIOR APPLICATION NUMBER: 60/264,649
8  PRIOR FILING DATE: 2001-01-26
9  NUMBER OF SEQ ID NOS: 23
10 SOFTWARE: Patentln Ver. 2.1
11 SEQ ID NO 10
12 LENGTH: 4985
13 TYPE: DNA
14 ORGANISM: Anopheles gambiae
15 US-10-056-405-10

```

Best Local Similarity	49.5%	Fixed: no	Indels	3	Gaps	3
Matches	333	Conservative	0	Mismatches	396	

[illegible]

Db	2770	ATATATATTAATAAATCCATATAAATAATTTATTTATGATATATATGATATAATGAT	2711
Qy	611	TAC-CTAAGATTTTACACCCAGATTGCTATPACCTTATPACATTCAAAGAAAGATCTTACA	669
Db	2710	AACATATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAAT	2651
Qy	670	AATACGCTCAACACTGCAGTGTCTTCCTTGGCCACTAGGTAGAGTTAGTGGAAACAATGCT	729
Db	2650	TATATTGTTATATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2591
Qy	730	GATGATGGAGTGTGATTTGCGGAATGCTCAATAACTATACAGATCCATATGCTCA	789
Db	2390	AATATATATAATATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	2531
Qy	790	ACTAAACACACTACTGATTAATCAAAATCCCTCAACTTCTTAATTCAGAGCAATGCGTGT	849
Db	2530	AGTAAATATTTTAATAATAATAGTAATATATATATATAATATAATATAATATAAT	2471
Qy	850	GCAACACATAGATCGATTCTCAATTGATGTCACAGCATAGATAATTAACAACCTTTCCAA	909
Db	2470	AATATATATAATATAATATAAATAAATAATATATATATATATAATATAAATAAATAAT	2411
Qy	910	TTAGATGAAAAATTTGTTTATCCAGAAATGACGTGGTCTGGAAGAATAAATAATTACA	969
Db	2410	AATATATATAATATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAT	2351
Qy	970	AGATTAGTACT 981	
Db	2350	AAATTCGTGATT 2339	

```

RESULT 7
US-10-094-240-10/c
Sequence 10, Application US/10094240
Publication No. US20030082637A1
GENERAL INFORMATION:
APPLICANT: ZWIEBEL, LAURENCE J.
TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
FILE REFERENCE: IN289
CURRENT APPLICATION NUMBER: US/10/094,240
CURRENT FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: 10/056,405
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: 60/264,649
PRIOR FILING DATE: 2001-01-26
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 4985
TYPE: DNA
ORGANISM: Anopheles gambiae
US-10-094-240-10

```

Query Match	2.0%;	Score 62.6;	DB 14;	Length 4985;
Best Local Similarity	45.5%;	Pred. No. 0.015;		
Matches 333;	Conservative 0;	Mismatches 396;	Indels 3;	Gaps 3;

Oy		253	TCCAAATTAGTACTTTTACGTAACCANTTAATGTGGTTAGATCTAAGTGATGGTTGGT	312
Dd		3070	ATTAAAAAAAACAATAATTAATPAAACANPAATAAGAATAACCAACAACAATAATAMAGA	3011
Oy		313	GGATACCAAAGTAAACAATAAGTTCA-GATTACACAACATGCAGAAATAGATTGGATCA	371
Dd		3010	CACACACACACAACATATATAGATTAATTAATTAACCATATGCCATAATTAAGATAATRA	2951
Oy		372	AAGCAAAACAAGACATATTATYGTCCTGTGGTTAATGATGAAGCATAAGTTCA-TTAA	430
Dd		2950	TAACAAATAGCAATAATATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2891
Oy		431	AAGAATAATACACCAATCTGAATGGTATTAAGAAACAACAATTCTAAAGTTTGAA	490
Dd		2890	AAAAAAAAATATTATTAATAGATAGTAAATAATAATTAATTAATTAATTAATTAATTAATTA	2831

Oy	491	TTGGTGGCTGGAATTAACGACCTGACGTAATTAGATTTACGATGATGGAGCTAAATTTA	550
Db	2830	ATAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2771
Oy	551	ATTTTACAAACCAACCTCAAGGTGAAATTTGTTAATGACTTCATTTTATGATGCCCAATCT	610
Db	2770	ATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2711
Oy	611	TAC-CTAAGAATTTTACACCGAGATTGGTATTAACCTTATACATTCAACAGAAAGTCTTACA	669
Db	2710	AACCATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2651
Oy	670	AATGACGTCAACACTGCAGTTGTTCTTGCCAGTAGGTAGAGTTAGTGGAAACAATGCT	729
Db	2650	TATTTATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2591
Oy	730	GATCATGGAGTGTGATTGTGGGAATGGTCAAAATTAACATTAACAGATCCATTTGCCAA	789
Db	2590	AATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2531
Oy	790	ACTAAACACACTACTGATTAATCAAAATCCCTCAACTTTTAAATTCAGAGCAATGCGCTGT	849
Db	2530	AGTAATATATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	2471
Oy	850	GCAACAATAGATAGATTCTCAATTTGATGATGTCAAGCTAGATAATTAACAATCTTTCCAA	909
Db	2470	AATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2411
Oy	910	TTAGATGAAAAATTTGTTATCCGAATGAGCTGGTCTGAGAGCAATAAAAATTTTCA	969
Db	2410	AATATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	2351
Oy	970	AGATTAGCTACT 981	
Db	2350	AAATTCGTGATT 2339	

```

RESULT 8
US-10-311-455-1798/C
; Sequence 1798, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Det
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1798
; LENGTH: 8771
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1798

```

	Query Match	1.9%	Score 60.8	DB 12	Length 8771
	Best Local Similarity	46.3%	Pred. NO. 0.047		
	Matches 200	Conservative 0	Mismatches 233	Indels 0	Gaps 0
Qy	..256	AAATTAGACTTTTACTAAACCATATATGCTTTAGATCTAGATGATGTTGGTGGGA	315		
Db	5053	ATATCTATACATCATCAATAAAATTTATATCAACGTTTAAAAAAAAAAAAACCTTTAACTTA	4994		

Query Match	1.9%	Score 60.2	DB 12	Length 3489
Best Local Similarity	61.6%	Pred. NO. 0	04	
Matches 114	Conservative	0	Mismatches 68	Indels 3
				Gaps 1
QY	2998	CCTGCTGGTTTAGGTCGTGTAATCTGTGATAGAAACCTGCTGCTGCTGCTAAACCT	3057	
Db	2236	CATCTCGCGCTGCTGCTCATCTGCTGCTGCTGCTCATCTGCTGCTGCTGCTATCTT	2177	
QY	3058	GCTGCTCCAGCAAAACCATCTGCACCAAAAGTAGCTACCAAGCTAAACCACTGGGG-	3115	
Db	2176	GCTGCTGCTGCTCATCTCTGCTGCTGCTCATCTCGTGTGCTGCTCATCTGCTGCTCT	2117	
QY	3116	-CTAAATCTGTCGGCCTACAAAACCAACTGCTCTTAAGCGACGCTGCTCAAAACCAAC	3174	
Db	2116	GCTCATCTGCTGCTCTGCTGCTCATCTGCTGCTGCTCATCTCTGCTGCTCATCTGCT	2057	
QY	3175	GCTCC 3179		
Db	2056	GCTGC 2052		

```
/ CURRENT FILE NO.: 0000  
/ NUMBER OF SEQ ID NOS : 2  
/  
/ SEQ ID NO 1  
/ 000000
```

```

RESULT 12
US-10-349-680-148
: Sequence 148, Application US/10349680
: Publication NO. US20030176654A1
: GENERAL INFORMATION:
: APPLICANT: Cassell, Gall
: APPLICANT: Chen, Ellison
: APPLICANT: Glass, Jennifer
: APPLICANT: Glass, John
: APPLICANT: Heiner, Cheryl
: APPLICANT: Leikowitz, Elliot
: TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
: TITLE OF INVENTION: UREALYTICUM
: FILE REFERENCE: UAB-13403/22
: CURRENT APPLICATION NUMBER: US/10/349,680
: CURRENT FILING DATE: 2003-01-23
: PRIOR APPLICATION NUMBER: US 09/601,198
: PRIOR FILING DATE: 2000-12-08
: PRIOR APPLICATION NUMBER: PCT/US99/01972
: PRIOR FILING DATE: 1999-01-29
: PRIOR APPLICATION NUMBER: US 60/073,189
: PRIOR FILING DATE: 1998-01-30
: NUMBER OF SEQ ID NOS: 181
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 148
: LENGTH: 3057
: TYPE: DNA
: ORGANISM: Ureaplasma urealyticum
US-10-349-680-148

```

```

RESULT 13
US-10-091-007-55
? Sequence 55, Application US/10091007
? Publication NO. US20030170782A1
? GENERAL INFORMATION:
? APPLICANT: Microbial Technics limited
? APPLICANT: Le Page, Richard W F Wells, Jeremy M
? APPLICANT: Hanniffy, Sean B
? TITLE OF INVENTION: Proteins
? FILE REFERENCE: PMC/P21978WO
? CURRENT APPLICATION NUMBER: US/10/091.007
? CURRENT FILING DATE: 2002-03-06
? PRIOR APPLICATION NUMBER: GB 9291125.2
? PRIOR FILING DATE: 1999-09-07
? NUMBER OF SEQ ID NOS: 276
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 55
? LENGTH: 2451
? TYPE: DNA
? ORGANISM: Streptococcus agalactiae
US-10-091-007-55

```

	Query Match	Best Local Similarity	1.6%	Score 51.6;	DB 12;	Length 2451;
	Matches 189;	Conservative	0;	Mismatches 229;	Indels 0;	Gaps 0;
QY	1184	TATCTGATCTAACGCGTTTGGATCACTAAACAGATGATATTGACCTTCACGACTG	1244			
DB	1016	TTTATTAATTTCCACGATACATTTATTTCGCCCGATTAAGCCTTAATAATCATATCAAG	1077			
QY	1244	ATAACGCTACACAACAACGACGAACGACAAACGCTGATACATCTAGTGGTTCACAG	1303			
DB	1076	AAAAGGATTTTGTCGGAAAGCATATCAAGTTTGTGAGAGAAACATTTTGAGAGTTTGGTA	1133			

	Query Match	1.6%;	Score 50.8;	DB 14;	Length 2017;	
	Best Local Similarity	41.8%;	Pred. No.2.5;	532;	Indels	6; Gaps
	Matches	386;	Conservative	0;	Mismatches	1;
OY	207 TAGAGATGTTGGTGTACCTTTATTACAGCAGCACTAATGGAACAATTAATAATTTGATG	266				
Dd	82 TAAAAATGATATGATGAAAAAAAGAAAAGAATAATTAATATATACAAAGAGATATAT	141				
OY	267 TTTTACTAACAATTATATGTTTGATGCTAACAGATGATGTGGTGGATACAAAGTAA	326				
Dd	142 AAAAAATGCCAAATGTCATATTAATTAATTAATTAACAAATTCCTTAAGATATGCACAAA	201				
OY	327 ACNAATAGTTGGCATTCACACACTGCAGAAATAGATTGGATCCAAAGCAAACAAGAC	386				
Dd	202 TGCACTACATTAGATGAACACATATAATAGGTAAACCTTGAAGGTTATGTTATATGCT	261				
OY	387 ATATTATGCTCTGGTTAATGATGAGAGCTACGTTCACTTTAAAAAGAAATTAATCTA	446				
Dd	262 TAATAAAAAATGCGCTGTATTAAATTTAAAAAATTTAATTAATATATGTTGTTAATTAA	321				
OY	447 CTCAAATAGAAATGGTAAATGAACAACAACAAATCTTAAGTTGTAAATGSGTGTGTAA	506				
Dd	322 AAACAAGCAAAATTTAGTGATGACATTCAGATTCAGATATTTTTTTTGA AAAAGATCA	381				
OY	507 TCAGCTCACGTAATTAGATTACAGATGATGGGCACTAAATTAATTTACAAACCAAC	566				
Dd	382 ACCCGTTCAGATGAATAATCTGTGTATTATACAAAGAAAGAAATATATTATTCYAGAAA	441				
OY	567 TCAGGTGAATTTGTATGACTTCTATTTTAGATGCGCAATCTTACTAAAGATTTCA	626				
Dd	442 TATTATTAATATATGAAAAATATAAATTAAGTAAAGAGAAATATTCAAAAGGATTAAT	501				

[illegible]

Db 1560 TAATAAATAATGCTGCGTATTAATAAATAATTAATAATATGTTGTTATTTAA 1619
QY 447 CTCAATAGATTTGGTAAATAGAAACAACAATTCTAGTTGTAATTTGGTGTGTTAA 506
Db 1620 AACCAAGCAAAATTTAGTGATGACATGAGATATGAATGATTTTGGAAAAAGATCA 1679
QY 507 TCCAGCTCAGCTAATTTGATTTACTGATGAGTGGACTAATTTAATTTTACAAACCAAC 566
Db 1680 ACCGGTTCAATGTAATAATTTGTTGTTATTAATACAAAGAAAGAAATTTATTTCTAGAAA 1739
QY 567 TCAAGGTGAATTTGTTAATGATTTTATGATGCGCAATCTTACCTAAGATTTTCA 626
Db 1740 TATTATTAATTAATGAATAATTAATAATTAAGTAAGGAATAATTTCAAGGATTAAT 1799
QY 627 CCCAGATTTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 686
Db 1800 AACCAAGTATGATGCTGCTGTTTATTAAGTTTAAAAATGGAAGCACTGGATA 1859
QY 687 AGTTGTCCTTGGCAGTAGTAGTGAAGTGAACAAATGCTGATGAGGATGTTGA 746
Db 1860 TTTACATTAATCAAAATTTGTTGTTATGATTAAGAAAGAAATTA-----TGA 1913
QY 747 TTGTGGGAATGGTCAATACTAATACAGATCCTATGCTCAAACTAATAACCACTACTGA 806
Db 1914 TAATCAAAATTTATGATTAATCAAAATATGATTAATCAAAATTTATGATTAATCAAAATTTATGA 1973
QY 807 TAATCAAAATCTTCAACTTTTAATTCAGAGCAATGCTGTGCAAAACAATAGATACGA 866
Db 1974 TAATCAAAATTTAATTAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATGA 2033
QY 867 TTCATTAATGAATGTCAGCATGATGAATTAATAACATCTTCCAAATTAGATGAATAATTTGT 926
Db 2034 TAATCAAAATTTAATTAATTAATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATGA 2093
QY 927 TTATCCAGAAATGAGTCTGTAAGAAATTAATAATTTTACAGATTTAGCTACTGGAAG 986
Db 2094 TAATTTCAAAATGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 2153
QY 987 TTGGCCAGCAAGCAAGATATGATTTCTGACATACCGGGACTCCACAGTTACTTT 1046
Db 2154 TATTAATTAAGATTTGAATAATTTAGAACCTCCACAAATGATTAATTAATTAATTAATTT 2213
QY 1047 AAAAGAGATTTCAATTAATTTTCAAGCTTACTTAACCTCAAGTTAATTTCTTATC 1106
Db 2214 GACAAATACCAAGAGATCTAAACATTTTAAGAAATTTTCAATTAATTTGAATGTTTAA 2273
QY 1107 ATTCAATGCTGATAGTATTTATAT 1130
Db 2274 AGAAATGAGATATTAATAATAT 2297

Search completed: October 10, 2003, 16:32:18
Job time : 815 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 06:22:40 : Search time 11360 Seconds

(without alignments)
11484.225 Million cell updates/sec

Title: US-09-901-572a-2

Perfect score: 3189
Sequence: 1 atgaatattctcaaaaact.....caacgcctcccaagaataa 3189

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

```

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3189	100.0	3189	6	AX665165	AX665165 Sequence
2	3145.8	98.6	3189	1	AB023292	AB023292 Mycoplasma
3	3144.2	98.6	301042	1	AE016967	AE016967 Mycoplasma
4	3143.4	98.6	8354	1	AF214004	AF214004 Mycoplasma
5	3143.4	98.6	10651	6	AX113685	AX113685 Sequence
6	3057.8	95.9	3129	6	AX665242	AX665242 Sequence
7	1052.4	33.0	7141	1	AF083976	AF083976 Mycoplasma
8	993.2	31.1	1131	1	AB033210	AB033210 Mycoplasma
9	924.8	29.0	1128	1	AB033211	AB033211 Mycoplasma
10	124.6	3.9	8760	1	MYCWP	M31431 M.genitaliu
11	124.6	3.9	15787	1	U39698	U39698 Mycoplasma
12	124.6	3.9	110000	6	AR30198_2	Continuation (3 of
13	102.2	3.2	9691	1	MYCATP	M21519 M.pneumonia
14	102.2	3.2	16876	1	AE000002	AE000002 Mycoplasma
15	93	2.9	1618	1	MG034970	U34970 Mycoplasma
16	80.8	2.5	105682	3	AC116957_3	Continuation (4 of
17	78.4	2.5	4695	1	MGU34842	U34842 Mycoplasma
18	76.2	2.4	145265	2	AC007926	AC007926 Trypanoso
19	75.2	2.4	155204	2	AC007926	AC007926 Trypanoso
20	75.2	2.4	213544	2	BX510939	BX510939 Danio rer
21	73.6	2.3	151343	5	AL929250	AL929250 Zebrafish
22	72.4	2.3	94434	5	AL929250	AL929250 Zebrafish
23	71.8	2.3	64707	3	AC115607	AC115607 Dictyoste
24	71.8	2.3	225581	2	BX537105	BX537105 Danio rer
25	71.4	2.2	203138	2	BX530070	BX530070 Danio rer
26	70.8	2.2	3502	1	AY212515	AY212515 Mycoplasma
27	69.2	2.2	148871	2	AL929469	AL929469 Danio rer
28	68.8	2.2	266544	3	AC116956	AC116956 Dictyoste
29	68.6	2.2	151802	3	AC114263	AC114263 Dictyoste
30	67.6	2.1	82139	3	AC115684	AC115684 Dictyoste
31	67.4	2.1	12463	3	AC115578	AC115578 Dictyoste
32	67.4	2.1	168244	2	BX537273	BX537273 Danio rer
33	66.6	2.1	111882	3	AC115612	AC115612 Dictyoste
34	66.4	2.1	110000	3	AC116984_3	Continuation (4 of
35	66.4	2.1	180108	2	BX511115	BX511115 Danio rer
36	66	2.1	176267	2	AL929105	AL929105 Danio rer
37	65.8	2.1	250029	3	AE014820	AE014820 Plasmodiu
38	65.4	2.1	169478	2	AC139280	AC139280 Homo sapi
39	65.4	2.1	195532	2	BX004760	BX004760 Danio rer
40	65.4	2.1	250195	3	AE014831	AE014831 Plasmodiu
41	65.2	2.0	158266	2	BX323881	BX323881 Danio rer
42	65.2	2.0	178702	2	BX321884	BX321884 Danio rer
43	65.2	2.0	185107	2	AC118513	AC118513 Rattus no
44	65.2	2.0	243570	2	AC125757	AC125757 Rattus no
45	65	2.0	166583	2	AC137124	AC137124 Mus muscu

ALIGNMENTS

```

RESULT 1
LOCUS AX665165 3189 bp DNA linear PAT 26-MAR-2003
DEFINITION Sequence 2 from Patent EP1275716.
ACCESSION AX665165
VERSION AX665165.1 GI:29290295
KEYWORDS
SOURCE
ORGANISM Mycoplasma gallisepticum
Bacteri; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
REFERENCE 1
AUTHORS Okuda,T., Saito,S., Dorsey,K.M. and Tsuzaki,Y.
TITLE Modified dna molecule, recombinant containing the same thing, and
uses thereof
JOURNAL Patent: EP 1275716-A 2 15-JAN-2003;

```


OY 1981 ACCTTCTTAACCTAAATAGACCAATCCAAACGGCTAGAAATGATTCGTCACCAACA 2040
 |||||
 DB 1981 ACCTTCTTAACCTAAATAGACCAATCCAAACGGCTAGAAATGATTCGTCACCAACA 2040
 OY 2041 TACTTACATCAACAATTTGGATTACCTAGAACATCTGATTTACCAACCAACACCATTC 2100
 |||||
 DB 2041 TACTTACATCAACAATTTGGATTACCTAGAACATCTGATTTACCAACCAACACCATTC 2100
 OY 2101 GGAACAACGACCAAGTTATTTTCAGTATTCACCTGGTATGATTCATTCATCAATTAAGAAAT 2160
 |||||
 DB 2101 GGAACAACGACCAAGTTATTTTCAGTATTCACCTGGTATGATTCATTCATCAATTAAGAAAT 2160
 OY 2161 ATTAGAACATCTCTCCCTGGTAAACAGTATGATTCCTTATTCACAAATGAATAATAT 2220
 |||||
 DB 2161 ATTAGAACATCTCTCCCTGGTAAACAGTATGATTCCTTATTCACAAATGAATAATAT 2220
 OY 2221 AATATGATTTTATTAACATTAAAGATTAGCTGACTCAAGTACCCGTGATCGTCAAGCTCA 2280
 |||||
 DB 2221 AATATGATTTTATTAACATTAAAGATTAGCTGACTCAAGTACCCGTGATCGTCAAGCTCA 2280
 OY 2281 TTCAGTCCAAAGTTTATTAAGCTTAATGAATTTGGTATATCTTATCTTATTAAGAC 2340
 |||||
 DB 2281 TTCAGTCCAAAGTTTATTAAGCTTAATGAATTTGGTATATCTTATCTTATTAAGAC 2340
 OY 2341 AATTCATCTTATACAGTAAATCTGCTGGTAAATGTTGATGTTCTTCATCAAAACCTGGT 2400
 |||||
 DB 2341 AATTCATCTTATACAGTAAATCTGCTGGTAAATGTTGATGTTCTTCATCAAAACCTGGT 2400
 OY 2401 TCTCTGATCATATACCTGCTGATTAATACATTTAATCAGACTTATCTATATTCGTTT 2460
 |||||
 DB 2401 TCTCTGATCATATACCTGCTGATTAATACATTTAATCAGACTTATCTATATTCGTTT 2460
 OY 2461 GAAGGTTCTGGTCAATATCATCTGATTTCTGGGGAACAATCCAAATTCAAACCCGAT 2520
 |||||
 DB 2461 GAAGGTTCTGGTCAATATCATCTGATTTCTGGGGAACAATCCAAATTCAAACCCGAT 2520
 OY 2521 GAGTACTTAATTCAAATTTGGTTCATCTAGTCAAGTGGCTAGAATCTTCCTACCAACCA 2580
 |||||
 DB 2521 GAGTACTTAATTCAAATTTGGTTCATCTAGTCAAGTGGCTAGAATCTTCCTACCAACCA 2580
 OY 2581 AACTTCTTAACAGTTTACTTGAATTCATCTGCTGAATGCTGATTAAGTACTACCTGTA 2640
 |||||
 DB 2581 AACTTCTTAACAGTTTACTTGAATTCATCTGCTGAATGCTGATTAAGTACTACCTGTA 2640
 OY 2641 GTGGTGTGATCTGATGTAATTTTAAACAACCAACCTACCTTAAAGTTCAGATCCAA 2700
 |||||
 DB 2641 GTGGTGTGATCTGATGTAATTTTAAACAACCAACCTACCTTAAAGTTCAGATCCAA 2700
 OY 2701 TACTTATGATGTAATGATTAATTAAGAACAAATTAATTAAGTAACTTCTCT 2760
 |||||
 DB 2701 TACTTATGATGTAATGATTAATTAAGAACAAATTAATTAAGTAACTTCTCT 2760
 OY 2761 TATTAACAATTTGGGCTTACCTTCATGATGTAAGTCCCTACAGCAATTTGGTATCATTA 2820
 |||||
 DB 2761 TATTAACAATTTGGGCTTACCTTCATGATGTAAGTCCCTACAGCAATTTGGTATCATTA 2820
 OY 2821 GGTATCTCTGCAATATGATCATCTTAGGATTAGCATAGGATTCCTTTAAGAGCTCAA 2880
 |||||
 DB 2821 GGTATCTCTGCAATATGATCATCTTAGGATTAGCATAGGATTCCTTTAAGAGCTCAA 2880
 OY 2881 AGAAAAATTACAGACAAAGGTTTCAAAACAACTTCAAAAAAGTTGATACCTTGCT 2940
 |||||
 DB 2881 AGAAAAATTACAGACAAAGGTTTCAAAACAACTTCAAAAAAGTTGATACCTTGCT 2940
 OY 2941 GCTGTTGTTTCAAGTATTAACAAAGATTTATCCCAACCTGCTAAGTAAAGAAAAACCT 3000
 |||||
 DB 2941 GCTGTTGTTTCAAGTATTAACAAAGATTTATCCCAACCTGCTAAGTAAAGAAAAACCT 3000
 OY 3001 GGTGTTGTTTCAAGTATTAACAAAGATTTATCCCAACCTGCTAAGTAAAGAAAAACCT 3060
 |||||
 DB 3001 GGTGTTGTTTCAAGTATTAACAAAGATTTATCCCAACCTGCTAAGTAAAGAAAAACCT 3060
 OY 3061 GCTCAGCTAAACCATCTGACCCAAAGCTAGCTCAACGCTTAAACCAACTGGGCTAAA 3120

DB 3061 GCTCAGCTAAACCATCTGACCAAAAGCTACTACCAAGCTTAACCAACTGGGCTAAA 3120
 |||||
 OY 3121 TCTGTCGCGCTTACAAACCAACTGCTCTTAAGCAGCTGCTCCAAACCAACCGCTCCC 3180
 |||||
 DB 3121 TCTGTCGCGCTTACAAACCAACTGCTCTTAAGCAGCTGCTCCAAACCAACCGCTCCC 3180
 OY 3181 AAGAATAA 3189
 |||||
 DB 3181 AAGAATAA 3189
 |||||
 RESULT 2
 AB023292
 LOCUS
 DEFINITION
 AB023292 3189 bp DNA linear BCT 27-MAY-2000
 Mycoplasma gallisepticum gene for 120kDa membrane protein, complete cds.
 ACCESSION
 AB023292
 VERSION
 AB023292.1 GI:4240363
 KEYWORDS
 120kDa membrane protein.
 SOURCE
 Mycoplasma gallisepticum
 ORGANISM
 Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 REFERENCE
 1 (sites)
 Yoshida,S., Fujisawa,A., Tsuzaki,Y. and Saitoh,S.
 Identification and expression of a Mycoplasma gallisepticum surface antigen recognized by a monoclonal antibody capable of inhibiting both growth and metabolism
 Infect. Immun. 68 (6), 3186-3192 (2000)
 JOURNAL
 MEDLINE
 20278096
 PUBMED
 10816462
 REFERENCE
 2 (bases 1 to 3189)
 Yoshida,S.
 Direct Submission
 Submitted (02-FEB-1999) Shigeto Yoshida, Jichi Medical School,
 Department of Medical Zoology, Yakushiji 331-1,
 Minamikawachimachi, Tochigi 329-0498, Japan
 (E-mail:shigeto@jichi.ac.jp, Tel:81-285-58-7339,
 Fax:81-285-44-6489)
 FEATURES
 source
 location/Qualifiers
 1..3189
 /organism="Mycoplasma gallisepticum"
 /mol_type="genomic DNA"
 /strain="R"
 /db_xref="taxon:2096"
 1..3189
 /gene="M11"
 1..3189
 /gene="M11"
 /codon_start=-1
 /transl_table=4
 /product="120kDa membrane protein"
 /protein_id="BA74842.1"
 /db_xref="GI:4240364"
 /translation="MNISKKLSYLLIGLAVFAGALGSAPFGKSDKSNDFTOVLNO
 ARTLDANSVRLAGLQNGSLFNTVLDVVDNITANGTIIKIDSTPKLYGLDLSDD
 FAGYVKQIVSYTTSRNFDRORRAYALVLEBANHYLRINTNSRKNRNNNS
 KRVIGQVNDPAHVIVPTDCTKPNFKQKQGTVDFTDAPILPDLIPDMYNIAYO
 KRIIPDVATFAVPMVGVSGTNADGQFEDNGCOITFTDPIAOTKTTIDMONSTF
 NSGAMPGANRRYDOLNVRHKRTSFQLEKRVYFPMTSSEENKNTRIATLSLPSF
 RYWIIDLPQTPVTLKEDSVNYSRLVNSVSLISFISGSIYIFGSELSIMYSFP
 TRLSRDTALNQKTDIEASSTIDNGTNTTADTSSTGSGTACGANTNTSQVSN
 PTLNTRYRSFGIDSKPSASANKIDETNADNDNVLEARIYAEIRLIONEIPITNAGNFR
 NTIGVGFSTSGSRVYLRSYNDQPTNCPDPLVRCYLQCRGCTGCTPGYATKL
 LANSPIVDVDAARVGETHOFRTSLTYFVWGGYLTBEGARFSNTPTIRAGDTPES
 RSTFQSGYSDNTYEXYIOSVLAGEGILNNLVGKASSFLNSNRPNGLEMLAATYLS
 RSQIGLARTSGLPNOQPFQTHQVLSVSGDQFSSIKNRTIIFPGNQLWYFLFTNEN
 KSEVYLRILADSNDPSSSEPTSLIDVNEIGVILPLDNSFYTVNAGNYALFSSN
 PCSPGSTVAVTFNPNOLSDIAEFGSAGATSPDWGQIQRPDEYILIONGFTSOVAFN
 VTNOSFLNSVTFPTPANAGTNRKYVVDPPGNTLNQLPKIQOYQIDGKYXAKIKLN
 NLVTFSYNNEPALPSNVPEFAIGSTGLIAIKHITLDAIGIFLRAOKRIQDDGFKTF
 KKVDTLTAAVGSVYKRIITQTAHVKKRPALAGSKSGDKRPAANAAPAPAPSPAKRA
 SSPAKPTAPKSGAPKPTAPKPAAPKPTAPKE"

BASE COUNT 1129 a 571 c 525 g 964 t
ORIGIN
Query Match 98.6%; Score 3145.8; DB 1; Length 3189;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 3162; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 1 ATGATATTTCTAAAAAACTTAAAGTTATATACATTGATGATGCGATTAGCTGTATTGGA 60
DB 1 ATGATATTTCTAAAAAACTTAAAGTTATATACATTGATGATGCGATTAGCTGTATTGGA 60
QY 61 GCTCTGTGTTCTGCAAGCTTGGCTTTAAGCAATCAGATTAAGAGTAACGATACACGCA 120
DB 61 GCTCTGTGTTCTGCAAGCTTGGCTTTAAGCAATCAGATTAAGAGTAACGATACACGCA 120
QY 121 TTAGTTAATCAAGCAAGAAAGCTAGATGCTAATTCCTGTTAGACTTGCAGGCTTGGACAA 180
DB 121 TTAGTTAATCAAGCAAGAAAGCTAGATGCTAATTCCTGTTAGACTTGCAGGCTTGGACAA 180
QY 181 AATGTTCTGTTCTCAATACAGTTCTTGAAGATGTTGATGATACCTTTATACAGCAGCT 240
DB 181 AATGTTCTGTTCTCAATACAGTTCTTGAAGATGTTGATGATACCTTTATACAGCAGCT 240
QY 241 AATGGAACAATTTCAAAATTAGATAGTTTACTTAAACCAATTAATGCTTTAGATCTAAGT 300
DB 241 AATGGAACAATTTCAAAATTAGATAGTTTACTTAAACCAATTAATGCTTTAGATCTAAGT 300
QY 301 GATGATTTGCTGATACAAAGTAAAAAACAATAGTTTCAGATTACCAACTAGCAGAAAT 360
DB 301 GATGATTTGCTGATACAAAGTAAAAAACAATAGTTTCAGATTACCAACTAGCAGAAAT 360
QY 361 AGATTGATCAAGCAAGCAAGAGCATATTATGCTGTGTGTTAATGATGAAGCTAAC 420
DB 361 AGATTGATCAAGCAAGCAAGAGCATATTATGCTGTGTGTTAATGATGAAGCTAAC 420
QY 421 GTTCATTAAAAAAGATTAATACACTCAAAATGATGATGATGATGATGATGATGATGAT 480
DB 421 GTTCATTAAAAAAGATTAATACACTCAAAATGATGATGATGATGATGATGATGATGAT 480
QY 481 AAGTTTGAATTTGCTGTTGATTAATCCAGCTACGATTAATGATGATGATGATGATGATGAT 540
DB 481 AAGTTTGAATTTGCTGTTGATTAATCCAGCTACGATTAATGATGATGATGATGATGATGAT 540
QY 541 ACTAAATTTAATTTTCAAAACCAACTCAAGTGAATTTGTTAATGACTTCAATTTAGAT 600
DB 541 ACTAAATTTAATTTTCAAAACCAACTCAAGTGAATTTGTTAATGACTTCAATTTAGAT 600
QY 601 GCGCCAAATCTTACCTAAAGATTTACCCAGATTTGATTAATCTTATACATTTCAAGAAAG 660
DB 601 GCGCCAAATCTTACCTAAAGATTTACCCAGATTTGATTAATCTTATACATTTCAAGAAAG 660
QY 661 ATCTTACCAAAATGAGTCAAGCTCAGTGTCTTGGCAGTAGAGTAGAGTAGAGTAGAGTAG 720
DB 661 ATCTTACCAAAATGAGTCAAGCTCAGTGTCTTGGCAGTAGAGTAGAGTAGAGTAGAGTAG 720
QY 721 ACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 721 ACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 AATGCTCAAACTAAACCACTACTGATTAATCAAAATCCCTCAACTTTAATTCAGAGACA 840
DB 781 AATGCTCAAACTAAACCACTACTGATTAATCAAAATCCCTCAACTTTAATTCAGAGACA 840
QY 841 ATGCTGTGTCAGCAACATAGATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 841 ATGCTGTGTCAGCAACATAGATAGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 901 TCTTTCCAAATTTAGTGAATAATTTGTTTATCCAGATTAAGTAAAGTAAAGTAAAGTAA 960
DB 901 TCTTTCCAAATTTAGTGAATAATTTGTTTATCCAGATTAAGTAAAGTAAAGTAAAGTAA 960
QY 961 AATATTAAGATAGATAGCTAGAGTGGTCCAGAGCAAGAAAGATTTGATTTCTTGAC 1020
DB 961 AATATTAAGATAGATAGCTAGAGTGGTCCAGAGCAAGAAAGATTTGATTTCTTGAC 1020

DB 961 AATATTAAGATAGCTAGAGTGGTCCAGAGCAAGAAAGATTTGATTTCTTGAC 1020
QY 1021 ATACCCGGAGCTCCAGCAAGTTACTTTAAAGAGATTCAATTAACGATTTTCAACAGTA 1080
DB 1021 ATACCCGGAGCTCCAGCAAGTTACTTTAAAGAGATTCAATTAACGATTTTCAACAGTA 1080
QY 1081 TACTTAAACCTAGTTAATCTTTTATCAATTCCTGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 TACTTAAACCTAGTTAATCTTTTATCAATTCCTGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 TCTGAATTAACATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 TCTGAATTAACATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 TTGATTCAGTTTAAACAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 TTGATTCAGTTTAAACAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 AACGGAACAAGCAACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 AACGGAACAAGCAACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 ACTACTAATCTTCAACAGTTTCTAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAA 1380
DB 1321 ACTACTAATCTTCAACAGTTTCTAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAA 1380
QY 1381 ATTGATAGTAAACCAACTTCTGCAACCAATTAATGATGATGATGATGATGATGATGATGATGAT 1440
DB 1381 ATTGATAGTAAACCAACTTCTGCAACCAATTAATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 GTTATTTAAGCAAGAAATTAATGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1441 GTTATTTAAGCAAGAAATTAATGCTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1501 ACTAATGCAAGAACTTTATCCAAACACATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1560
DB 1501 ACTAATGCAAGAACTTTATCCAAACACATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1560
QY 1561 TCAAGAT 1620
DB 1561 TCAAGAT 1620
QY 1621 CCTTTCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
DB 1621 CCTTTCTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 TACGGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
DB 1681 TACGGAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
QY 1741 GGTACTGAACCAATCAATTTAGAGAACTTCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
DB 1741 GGTACTGAACCAATCAATTTAGAGAACTTCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1800
QY 1801 CTAACTGAAGAAAGGCTAGAGATTTCTAATCTCCTAATTAATTAATTAATTAATTAATTAATTAAT 1860
DB 1801 CTAACTGAAGAAAGGCTAGAGATTTCTAATCTCCTAATTAATTAATTAATTAATTAATTAATTAAT 1860
QY 1861 ACACGGAAGAGCCGAGGATCTTCCAAATCTGCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1920
DB 1861 ACACGGAAGAGCCGAGGATCTTCCAAATCTGCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1920
QY 1921 CAATGCTTTTAAAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980
DB 1921 CAATGCTTTTAAAGATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1980
QY 1981 AGCTTTTAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040
DB 1981 AGCTTTTAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2040
QY 2041 TACTTAAAGATCAAAATTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100
DB 2041 TACTTAAAGATCAAAATTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2100

OY 2101 GGAGCACTCACCAAGTTATTTAGTATCACCCTGGTATCAGTTCTCATTAATAGAAAT 2160
 |||||
 Db 2101 GGAGCACTCACCAAGTTATTTAGTATCACCCTGGTATCAGTTCTCATTAATAGAAAT 2160
 |||||
 OY 2161 ATTAGAACATCTTCCCTGGTAACCAAGTATGCTATCTTATTCACAAATGAAT 2220
 |||||
 Db 2161 ATTAGAACATCTTCCCTGGTAACCAAGTATGCTATCTTATTCACAAATGAAT 2220
 |||||
 OY 2221 AATGAGTGTATATACATTAAGATAGTGTGACTCAAGTATGCTATGCTCAACCTGA 2280
 |||||
 Db 2221 AATGAGTGTATATACATTAAGATAGTGTGACTCAAGTATGCTATGCTCAACCTGA 2280
 |||||
 OY 2281 TTTCAGTCAACAAGTTTAAATGACGTTAAATGAAATGCTATATCTTATTTAGAC 2340
 |||||
 Db 2281 TTTCAGTCAACAAGTTTAAATGACGTTAAATGAAATGCTATATCTTATTTAGAC 2340
 |||||
 OY 2341 AATTCTATTTATACAGTAAATGCTGCTGTAATGTCATTTGTCATCAAAACCTGGT 2400
 |||||
 Db 2341 AATTCTATTTATACAGTAAATGCTGCTGTAATGTCATTTGTCATCAAAACCTGGT 2400
 |||||
 OY 2401 TCTCCTGATCATATATCTGCTGTAATATACATTTAATCAGACTATCTGATATGCTTT 2460
 |||||
 Db 2401 TCTCCTGATCATATATCTGCTGTAATATACATTTAATCAGACTATCTGATATGCTTT 2460
 |||||
 OY 2461 GAAGGTTCTGCTGCTAAGTATACATCTGATTTCTGGGGAACAATCCAAATCCGAT 2520
 |||||
 Db 2461 GAAGGTTCTGCTGCTAAGTATACATCTGATTTCTGGGGAACAATCCAAATCCGAT 2520
 |||||
 OY 2521 GAGTACTTAATTCAAAAGGTTGCTACTAGTCAAGTGGCTAGAAAGCTTCTTACAAACCA 2580
 |||||
 Db 2521 GAGTACTTAATTCAAAAGGTTGCTACTAGTCAAGTGGCTAGAAAGCTTCTTACAAACCA 2580
 |||||
 OY 2581 AGCTTCTTAACAGTTTATGACTGCTGCTGCTAATGCTGCTAATGCTAATGCTGTA 2640
 |||||
 Db 2581 AGCTTCTTAACAGTTTATGACTGCTGCTGCTAATGCTGCTAATGCTAATGCTGTA 2640
 |||||
 OY 2641 GTGGTGTATCTGATGCTGCTAATTTAACAAACCAAAACCTCTTAAAGTTCAGATCCAA 2700
 |||||
 Db 2641 GTGGTGTATCTGATGCTGCTAATTTAACAAACCAAAACCTCTTAAAGTTCAGATCCAA 2700
 |||||
 OY 2701 TACTTATGATGCTAATTTATGATGCTAAATTTAAAGAACAAATTAATTAAGTACATTTCTCT 2760
 |||||
 Db 2701 TACTTATGATGCTAATTTATGATGCTAAATTTAAAGAACAAATTAATTAAGTACATTTCTCT 2760
 |||||
 OY 2761 TATTAACACTTGGGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 |||||
 Db 2761 TATTAACACTTGGGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 |||||
 OY 2821 GGTATTTCTGCAATTTATGATCATCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
 |||||
 Db 2821 GGTATTTCTGCAATTTATGATCATCTTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880
 |||||
 OY 2881 AGAAAATTTACAAAGAGGTTCAAAGCAATTCAAAAAGTTGATACCTTGACTGCT 2940
 |||||
 Db 2881 AGAAAATTTACAAAGAGGTTCAAAGCAATTCAAAAAGTTGATACCTTGACTGCT 2940
 |||||
 OY 2941 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
 |||||
 Db 2941 GCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000
 |||||
 OY 3001 GCTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
 |||||
 Db 3001 GCTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3060
 |||||
 OY 3061 GCTCCAGCTAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3120
 |||||
 Db 3061 GCTCCAGCTAAACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3120
 |||||
 OY 3121 TCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
 |||||
 Db 3121 TCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
 |||||

OY 3181 AAGAATATA 3189
 |||||
 Db 3181 AAGAATATA 3189
 |||||
 RESULT 3
 AE016967
 LOCUS
 DEFINITION
 Mycoplasma gallisepticum strain R section 1 of 4 of the complete genome
 ACCESSION
 AE016967 AE015450
 VERSION
 AE016967.1 GI:31541048
 KEYWORDS
 SOURCE
 ORGANISM
 Mycoplasma gallisepticum R
 Bacteria; Filicutes; Mollicutes; Mycoplasmatellaceae; Mycoplasma.
 REFERENCE
 1 (bases 1 to 301042)
 Geary,S.J., Papazisi,L., Kutish,G., Gorton,T.S., Mahairas,G., Swartzell,S., Madan,A., Nguyen,D.K., Markham,P., Browning,G., Kamal,M. and Liao,X.
 The complete genome sequence of the avian pathogen Mycoplasma gallisepticum strain R
 Microbiology (2003) In press
 2 (bases 1 to 301042)
 Geary,S.J., Papazisi,L., Kutish,G., Mahairas,G., Swartzell,S., Madan,A., Nguyen,D.K., Gorton,T.S., Markham,P., Browning,G., Mustafa,K. and Liao,X.
 Direct Submission
 TITLE
 Submitted (17-OCT-2002) Department of Pathobiology and Veterinary Sciences, and Center of Excellence for Vaccine Research, The University of Connecticut, 61 North Eagleville Road U-89, Storrs, CT 06269-3089, USA
 JOURNAL
 FEATURES
 source
 location/Qualifiers
 1. 301042
 /organism="Mycoplasma gallisepticum R"
 /mol_type="genomic DNA"
 /strain="R"
 /db_xref="taxon:233150"
 1. 1889
 /note="putative oriC"
 complement(24. .32)
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma pulmonis consensus tttatcama"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 complement(36. .44)
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma pulmonis consensus tttatcama"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 114. .122
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 209. .217
 /note="oriC imperfect repeat"
 complement(254. .262)
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma mycoides consensus yttwMhamR"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 298. .306
 /note="putative DnaA-box; Mycoplasma spp. putative dnaA binding site; consensus sequence approach; Mycoplasma mycoides consensus yttwMhamR"
 /bound_molety="DnaA"
 /function="initiation of the chromosome replication"
 complement(401. .450)
 /note="oriC imperfect repeat"
 repeat_region

QY	121	TTAGTTAATCAAGCAAGACCGTACATGCTAAATTCCTGTTAGACTTCGAGCTCTTGACAA	180
Db	226727	TTAGTTAATCAAGCAAGAACCGCTAGATGCTAAATTCCTGTTAGACTTCGAGCTCTTGACAA	226786
QY	181	AATGGTCGTGTTCAATACAGTTCCTGTAGAGTGTGATGATTAACCTTTAATACACAGCT	240
Db	226787	AATGGTCGTGTTCAATACAGTTCCTGTAGAGTGTGATGATTAACCTTTAATACACAGCT	226846
QY	241	AATGGAACAATATCAAAATTAGATAGTATTTACTAAACATATATATGCTTTAGATCTAACT	300
Db	226847	AATGGAACAATATCAAAATTAGATAGTATTTACTAAACATATATATGCTTTAGATCTAACT	226906
QY	301	GATGATTTGTGCTGATACAAAGTAACAAATAGTTTCAGATTACACAATGACAGAAAT	360
Db	226907	GATGATTTGTGCTGATACAAAGTAACAAATAGTTTCAGATTACACAATGACAGAAAT	226966
QY	361	AGATTTGATCAAAAGCAACAAGACATATTATGCTCTGTGCTTAATGATGATAGCTAAC	420
Db	226967	AGATTTGATCAAAAGCAACAAGACATATTATGCTCTGTGCTTAATGATGATAGCTAAC	227026
QY	421	GTTCAATTAACAAAAGATTAATACATACATCAAAATAGAAATGGTAATAGAAACAACATTC	480
Db	227027	GTTCAATTAACAAAAGATTAATACATACATCAAAATAGAAATGGTAATAGAAACAACATTC	227086
QY	481	AAGTTTGTAAATTTGTGTGTTGATTAATCCAGCTCAAGTAATAGATTTACTGATGATGG	540
Db	227087	AAGTTTGTAAATTTGTGTGTTGATTAATCCAGCTCAAGTAATAGATTTACTGATGATGG	227146
QY	541	ACTAATATTAATTTTACAAACCAACTCAAGGTGAATTTGTTAATGACTTCATTTAGAT	600
Db	227147	ACTAATATTAATTTTACAAACCAACTCAAGGTGAATTTGTTAATGACTTCATTTAGAT	227206
QY	601	GGCGCAATCTTAACCTAAAGATTTTACACCAGATGTGTAATCTTATACCTTAACAAAGAAG	660
Db	227207	GGCGCAATCTTAACCTAAAGATTTTACACCAGATGTGTAATCTTATACCTTAACAAAGAAG	227266
QY	661	ATCTTACCAATGAGCTCAACACTGCTAGTGTGTTCTTGCCAGTAGAGTAGTATGATGA	720
Db	227267	ATCTTACCAATGAGCTCAACACTGCTAGTGTGTTCTTGCCAGTAGAGTAGTATGATGA	227326
QY	721	ACAAATGCTGATGATGGGATGTTTATTTGTGGAAATGTCAAATACTAATACAGATCCT	780
Db	227327	ACAAATGCTGATGATGGGATGTTTATTTGTGGAAATGTCAAATACTAATACAGATCCT	227386
QY	781	ATTTGCTCAACCTAAACACACTACGTATATCAAAATTCCTCAACTTTTATTCAGAGAGA	840
Db	227387	ATTTGCTCAACCTAAACACACTACGTATATCAAAATTCCTCAACTTTTATTCAGAGAGA	227446
QY	841	ATGCTCGTGTGCAACAAATAGATACATTTCTCAATTTGAATGTCAAGCATAGAAATTAACA	900
Db	227447	ATGCTCGTGTGCAACAAATAGATACATTTCTCAATTTGAATGTCAAGCATAGAAATTAACA	227506
QY	901	TCCTTCCAAATAGATGAATAAATTTGTTTATCCAGATGAGACGTGTTCTGAAGAAATAAA	960
Db	227507	TCCTTCCAAATAGATGAATAAATTTGTTTATCCAGATGAGACGTGTTCTGAAGAAATAAA	227566
QY	961	AATATTAACAAGTTTGTCTACTGAGGAGTTGCCAACCAACGAAGAATATGGAAATTCCTGAC	1020
Db	227567	AATATTAACAAGTTTGTCTACTGAGGAGTTGCCAACCAACGAAGAATATGGAAATTCCTGAC	227626
QY	1021	ATACCCGGAGACCCCAAGATTAATCTTTAAAGAAGATTCAGTTAACCTATTTTCAAGACTA	1080
Db	227627	ATACCCGGAGACCCCAAGATTAATCTTTAAAGAAGATTCAGTTAACCTATTTTCAAGACTA	227686
QY	1081	TACTTAAACCTCAGTTAATTCCTTATCATTTCAATTTGATAGTATTAATATTTTGGTACC	1140
Db	227687	TACTTAAACCTCAGTTAATTCCTTATCATTTCAATTTGATAGTATTAATATTTTGGTACC	227746
QY	1141	TCTGAATTAACCATTAATAGTACTATTTCAATTTCCCAACTAGATTAATCTGATCTAACCGCT	1200
Db	227747	TCTGAATTAACCATTAATAGTACTATTTCAATTTCCCAACTAGATTAATCTGATCTAACCGCT	227806

QY	1201	TTGATTCAGATTAAAAAGATATATATTGAAGCTTCAGACGACTGATTAACGGTACACACAA	1260
Db	227807	TTGAATCAAGTTAAAAACAGATATATTGAAGCTTCAGACGACTGATTAACGGTACACACAA	227866
QY	1261	AACGAAACAGCAACAGCTGATTCATCTAGCGTTCAACAGGTCCTGGAACAGGAAT	1320
Db	227867	AACGAAACAGCAACAGCTGATTCATCTAGCGTTCAACAGGTCCTGGAACAGGAAT	227926
QY	1321	ACTATAACACTTCTCAAAACAGTTTCTAATCCTACTTTAAATACCTATTCGTAGTTTGGAC	1380
Db	227927	ACTATAACACTTCTCAAAACAGTTTCTAATCCTACTTTAAATACCTATTCGTAGTTTGGAC	227986
QY	1381	ATTGATAGTAACCAACTTCTGCAACAAATAATAGATGAAGTAATGGCAGATCTTAAC	1440
Db	227987	ATTGATAGTAACCAACTTCTGCAACAAATAATAGATGAAGTAATGGCAGATCTTAAT	228046
QY	1441	GTTATTGAAGCAAGAATATATGCTGAATTCACATTAATAGTATTCAAAATGAATTCCAATA	1500
Db	228047	GTTATTGAAGCAAGAATATATGCTGAATTCACATTAATAGTATTCAAAATGAATTCCAATA	228106
QY	1501	ACTAATGCAGAAACCTTTATCCGAAACACAATGGTGGCTGGTTTACTTCAACAGGT	1560
Db	228107	ACTAATGCAGAAACCTTTATCCGAAACACAATGGTGGCTGGTTTACTTCAACAGGT	228166
QY	1561	TCACAGATGATTTTAAAGAGCTTCTATTAACGGTGATCAACGTCCAACTGSAACCTTCCAA	1620
Db	228167	TCACAGATGATTTTAAAGAGCTTCTATTAACGGTGATCAACGTCCAACTGSAACCTTCCAA	228226
QY	1621	CCTTCTCTATACGTAATTTGGTTATTTAGATACCAACAACTGAAACAGGAACCTTCTCG	1680
Db	228227	CCTTCTCTATACGTAATTTGGTTATTTAGATACCAACAACTGAAACAGGAACCTTCTCGA	228286
QY	1681	TACGGAACATATACGCTTTTAAACAAACAGCCCTTACGAGGTATTAATTCCTCCAAAGTA	1740
Db	228287	TACGGAACATATACGCTTTTAAACAAACAGCCCTTACGAGGTATTAATTCCTCCAAAGTA	228346
QY	1741	GGTACTGAAACCAATTCATTTAGAACAGCTTCATTAACTAACCTGTATAGGTGATAT	1800
Db	228347	GGTACTGAAACCAATTCATTTAGAACAGCTTCATTAACTAACCTGTATAGGTGATAT	228406
QY	1801	CTAATCTGAAGAAGGTCCTGAAGTTCCTTAATACCTCATATATTAGAGCACAGGTGAC	1860
Db	228407	CTAATCTGAAGAAGGTCCTGAAGTTCCTTAATACCTCATATATTAGAGCACAGGTGAC	228466
QY	1861	ACACAGAAAGCCGGAAGCATCTTCCAAATCGGCTATTCGATTAATCCTATGAGTACATT	1920
Db	228467	ACACAGAAAGCCGGAAGCATCTTCCAAATCGGCTATTCGATTAATCCTATGAGTACATT	228526
QY	1921	CAATCAGTTTGAAGATTTGATGGAATTTAGAAATTAACTTAAATGTTGGGTTAAAGCATCA	1980
Db	228527	CAATCAGTTTGAAGATTTGATGGAATTTAGAAATTAACTTAAATGTTGGGTTAAAGCATCA	228586
QY	1981	AGCTTCTTAACTCAAAATGACCAAAATCCAAAGGCTGAGAAATGATTTGCTGCAACAA	2040
Db	228587	AGCTTCTTAACTCAAAATGACCAAAATCCAAAGGCTGAGAAATGATTTGCTGCAACAA	228646
QY	2041	TACTTAAGATTCACAAATTTGGATTAGCTAGACATCGGATTTACCAACCAACCAACCTTC	2100
Db	228647	TACTTAAGATTCACAAATTTGGATTAGCTAGACATCGGATTTACCAACCAACCAACCTTC	228706
QY	2101	GGACAAACTCACCAAGTTATTTTCAGATACACTGGATGATCAGTTCTCATCAATTAAGAAAT	2160
Db	228707	GGACAAACTCACCAAGTTATTTTCAGATACACTGGATGATCAGTTCTCATCAATTAAGAAAT	228766
QY	2161	ATTAGAACAACTCTCCCTGGTAAACCGATTATGTTACTTCTTATTCACAAATGAAAATAT	2220
Db	228767	ATTAGAACAACTCTCCCTGGTAAACCGATTATGTTACTTCTTATTCACAAATGAAAATAT	228826
QY	2221	AAATCGAGTGTATATCATTAAGATTAAGTACGATCAAGTAACCTCATGCGTCAACCTCA	2280
Db	228827	AAATCGAGTGTATATCATTAAGATTAAGTACGATCAAGTAACCTCATGCGTCAACCTCA	228886
QY	2281	TTTCAGTCCAAACAGTTTAATGACGTTTAATGAATTTGTTGTAATCTTACCTTATTAGAC	2340

OY		2641	GTCGTGATCCGCATGATGAATTAAACCAACCACCAACTACCTTAATAAAGTTCAGATPCCA	2700
Dd		6418	GTGGTGATCCTCATGATGTAATTAAACAACCAACCACTACCTTAAGAAGTTTCAGATCCA	6477
OY		2701	TACCTAGATGATGAATATTATGATCTCTAAATTAAAGAACATATATTAGTACATTCCT	2760
OY		6478	TACTTAGATGATGAAGTATATATGATCTCTAAATTAAAGAACATATATTAGTACATTCCT	6537
Dd		2761	TATAACAACCTTGGGCCCTTACCTTCATGGGTAGTGCCTACAGCAATTGGTAGTACATTAA	2820
OY		6538	TATACACACTTTCCTGCTTACCTTCATGAGTAGTGCCCTACAGCAATTGGTAGTACATTAA	6597
Dd		2821	GGTATCTTCGCAATTATGATCATCTTAGATTAGTATCGTATTCCTTTAAGACTCAA	2880
OY		6598	GGTATCTTCGCAATTATGATCATCTTAGATTAGTATCGTATTCCTTTAAGAGCTCAA	6657
Dd		2881	AGAAAATTACACAGCAAGGGTTCCAAACACATTCCTAAAAAGGTGATACCTTACAGCT	2940
OY		6658	ACAAAATTACACAGCAAGGGTTCCAAACACATTCCTAAAAAGGTGATACCTTACAGCT	6717
Dd		2941	GCTGTGGTTCAGTTTACAGAAGATTTATACCAACACTGCTACAGCTTAAGAAAAACCT	3000
OY		6718	GCTGTGGTTCAGTTTACAGAAGATTTATACCAACACTGCTACAGCTTAAGAAAAACCT	6777
Dd		3001	GCTGCTTTAGAGTGTGCTTAATCTGTGATPAAGAAACCTGCTGCTGCTTAAACCTGCT	3060
OY		6778	GCTGCTTTAGAGTGTGCTTAATCTGTGATPAAGAAACCTGCTGCTGCTTAAACCTGCT	6837
Dd		3061	GCTCCAGCTTAAACCATCTGCGAACCAAAAGCTGCTCACCACTTAAACCAACCTGGCCTTAA	3120
OY		6838	GCTCCAGCTTAAACCATCTGCGAACCAAAAGCTGCTCACCACTTAAACCAACCTGGCCTTAA	6897
Dd		3121	TCTGTGCGCCTTACAAAACCAACTGCTCTCTTAAGCGACAGTCTGCCAAAACCAACCGCTCCC	3180
OY		6898	TCTGTGCGCCTTACAAAACCAACTGCTCTCTTAAGCGACAGTCTGCCAAAACCAACCGCTCCC	6957
Dd		3181	AAAGAAATTA 3189	
OY		6958	AAAGAAATTA 6966	
Dd				
RESULT 5				
AX113685				
LOCUS	AX113685	10651 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0127254.			PAT 01-MAY-2001
ACCESSION	AX113685			
VERSION	AX113685.1	GI:13939855		
KEYWORDS	. synthetic construct			
SOURCE	synthetic construct			
ORGANISM	artificial sequences.			
REFERENCE	1			
AUTHORS	Schwarzmann, F.			
TITLE	Gene transfer vectors for treating autoimmune diseases and diseases			
JOURNAL	with immunopathogenesis by therapy			
FEATURES	Patent: WO 0127254-A 1 19-APR-2001;			
SOURCE	Schwarzmann, Fritz (DE)			
	Location/Qualifiers			
	1..10651			
	/organism="synthetic construct"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:32630"			
	/note="Plasmid"			
BASE COUNT	2782 a 2618 c 2490 g 2759 t	2 others		
ORIGIN				
Query Match	98.6%; Score 3143.4; DB 6; Length 10651;			
Best Local Similarity	99.1%; Pred. No. 0;			
Matches 3159; Conservative	2; Mismatches	28; Indels	0; Gaps	0
OY	1 ATGAATATTTCTAAAAAACTTAAGCTTATACATTGATAGTGATGATTAAGTATTTGCA	60		
Dd	6 ATGAATATTTCTAAAAAACTTAAGCTTATACATTGATAGTGATGATTAAGTATTTGCA	65		

OY	61	GCTCTGGTTCGCAAGCTTTGGCTTTAAGCAATCAGATAAGACTAACGATTAACAGCAA	120
Db	66	GCTCTGGTTCGCAAGCTTTGGCTTTAAGCAATCAGATAAGACTAACGATTAACAGCAA	125
OY	121	TTACTTATATCAAGCAAGACGCTGATGCTAAATCTGTATACACTTGACAGCTCTTGACAA	180
Db	126	TTACTTATATCAAGCAAGACGCTGATGCTAAATCTGTATACACTTGACAGCTCTTGACAA	185
OY	181	AATGGTTCGTTGTTCAATACAGTTCCTTAGAGATGTGATGTAACCTTATATTAACAGAGCT	240
Db	186	AATGGTTCGTTGTTCAATACAGTTCCTTAGAGATGTGATGTAACCTTATATTAACAGAGCT	245
OY	241	AATGGAACAATATCAAAATTTGATGATGTTTACTAAACCATATATGTTTAGATCTAAGT	300
Db	246	AATGGAACAATATCAAAATTTGATGATGTTTACTAAACCATATATGTTTAGATCTAAGT	305
OY	301	GATGATGTGGTGGATACCAAAAGTAAACCAATAGTTCAGATTCACACAACACAGAGAAAT	360
Db	306	GATGATGTGGTGGATACCAAAAGTAAACCAATAGTTCAGATTCACACAACACAGAGAAAT	365
OY	361	AGATTTTGATCAAAACCAAGACGATATATGCTCTGTGGTTTAATGATGAAGCTTAAC	420
Db	366	AGATTTTGATCAAAACCAAGACGATATATGCTCTGTGGTTTAATGATGAAGCTTAAC	425
OY	421	GTCATTTTAAAAAAGTAATTAATACTAACCAATAGAAATGGTAATAGAAACACAATTC	480
Db	426	GTCATTTTAAAAAAGTAATTAATACTAACCAATAGAAATGGTAATAGAAACACAATTC	485
OY	481	AAGTTTGTATTTGGTGTGTGATTAATCCAGCTCAGCTAATTAAGATTTACTGATGATGG	540
Db	486	AAGTTTGTATTTGGTGTGTGATTAATCCAGCTCAGCTAATTAAGATTTACTGATGATGG	545
OY	541	ACTTAATTTAATTTTACCAACCAACCTCAAGTGAATGTAATGATCTTATATTTAGAT	600
Db	546	ACTTAATTTAATTTTACCAACCAACCTCAAGTGAATGTAATGATCTTATATTTAGAT	605
OY	601	GGGGCAATCTTACCTAAAGATTTACACCCAGATTTGCTTAATCTATACATTCCAAGAAAG	660
Db	606	GGGGCAATCTTACCTAAAGATTTACACCCAGATTTGCTTAATCTATACATTCCAAGAAAG	665
OY	661	ATCTTACCAATAGACGTCAACACGTGCAATGTTCTCTGGCCAGTAGTAGTAGTAGTGA	720
Db	666	ATCTTACCAATAGACGTCAACACGTGCAATGTTCTCTGGCCAGTAGTAGTAGTAGTGA	725
OY	721	ACAATAGCTGATGATGGGATGTTTGATGTGGGAATGGTCAAAATTAACAAATACAGATGCT	780
Db	726	ACAATAGCTGATGATGGGATGTTTGATGTGGGAATGGTCAAAATTAACAAATACAGATGCT	785
OY	781	ATTGCTCAACTAAAAACACACTACTGATATATCAAAATCCTTCAACTTTTAAATTCAGAGACA	840
Db	786	ATTGCTCAACTAAAAACACACTACTGATATATCAAAATCCTTCAACTTTTAAATTCAGAGACA	845
OY	841	ATGCGTGTGCAAAACAATAGATAGATTCCTCAATTTGATGTGCAAGCATAGAAATTAACCA	900
Db	846	ATGCGTGTGCAAAACAATAGATAGATTCCTCAATTTGATGTGCAAGCATAGAAATTAACCA	905
OY	901	TCTTTCCAAATAGATGAATAAATTTGTTTATTCAGAAATGAGACTGGTTCTGAAGAGAAATAA	960
Db	906	TCTTTCCAAATAGATGAATAAATTTGTTTATTCAGAAATGAGACTGGTTCTGAAGAGAAATAA	965
OY	961	AATATTTACAABATTTAGCTACTGGAAGTTTGGCAAGCAACGAAGATATTTGGATCTTAC	1020
Db	966	AATATTTACAABATTTAGCTACTGGAAGTTTGGCAAGCAACGAAGATATTTGGATCTTAC	1025
OY	1021	ATACCCGGAGATCCACAAGTACTACTTTAAAAAAGATTAGTTAAGTTTTCACAAGACTA	1080
Db	1026	ATACCCGGAGATCCACAAGTACTACTTTAAAAAAGATTAGTTAAGTTTTCACAAGACTA	1085
OY	1081	TACTTAAACTCAGTTAATTTCTTTATCATTCATTTGGTGATAGTATTTATATTTTTGGTACC	1140
Db	1086	TACTTAAACTCAGTTAATTTCTTTATCATTCATTTGGTGATAGTATTTATATTTTTGGTACT	1145

QY	1141	TCGTAATTACATCATTTATGGTACTATTCAATTCCTCCAACTAGATTATCTGATCTAACGGCT	1200
Db	1146	TCCTAAATTACCATCATTTATGTATGACTATTCAATTCCTCCAACTAGATTATCTGATCTAACGGCT	1205
QY	1201	TTGTAATCAAGTTAAACAGATGATATTGAAAGCTTCAAGCACTGATPAACGGTACCAACA	1260
Db	1206	TTGTAATCAAGTTAAACAGATGATATTGAAAGCTTCAAGCACTGATPAACGGTACCAACA	1265
QY	1261	AACGGACAAAGACAAACAGCTGATACATCTAGTGGTTCACAGGTGCTGGTAAACAGAAAT	1320
Db	1266	AACGGACAAAGACAAACAGCTGATACATCTAGTGGTTCACAGGTGCTGGTAAACAGAAAT	1325
QY	1321	ACTACTAACACTTCTCAACAGTTTCTATATCTTAAATTAATCTATTCGTAGTTTTGGA	1380
Db	1326	ACTACTAACACTTCTCAACAGTTTCTATATCTTAAATTAATCTATTCGTAGTTTTGGA	1385
QY	1381	ATTGATAGTAACCAACTCTCTGCAACCAAAATAGTGAACCTAAATTTGGGCGAGATCTTAC	1440
Db	1386	ATTGATAGTAACCAACTCTCTGCAACCAAAATAGTGAACCTAAATTTGGGCGAGATCTTAC	1445
QY	1441	GTTATTGAGCAAGAAATATATGCTGAATACAGATTAGTATTGATTCAAAATGAATTCATA	1500
Db	1446	GTTATTGAGCAAGAAATATATGCTGAATACAGATTAGTATTGATTCAAAATGAATTCATA	1505
QY	1501	ACTAATGCAAGAAACTTTATCCGAACAAATTTGGTGTGTTTGGTTTACTTCAACAGCT	1560
Db	1506	ACTAATGCAAGAAACTTTATCCGAACCAATTTGGTGTGTTTGGTTTACTTCAACAGCT	1565
QY	1561	TCAAGAGATGTTTATAGAGCTTCCTTATPAGGGTGATCAAGCTCCAACTGGAACCTGCCA	1620
Db	1566	TCAAGAGATGTTTATAGAGCTTCCTTATPAGGGTGATCAAGCTCCAACTGGAACCTGCCA	1625
QY	1621	CCTTTCTTATACGATTTTGGTTATTATAGATACCAACAACTAGAACAGAACTTTCGG	1680
Db	1626	CCTTCTTATACGATTTTGGTTATTATAGATACCAACAACTAGAACAGAACTTTCGA	1685
QY	1681	TACGGAACTATTAAGCTTTTAAACACAGCCCTTACGAGTATAGATTCTCCAGAGTA	1740
Db	1686	TACGGAACTATTAAGCTTTTAAACACAGCCCTTACGAGTATAGATTCTCCAGAGTA	1745
QY	1741	GGTACTGAACCAATCAATTTAGAAAGACTTCAATACATCCCGGTATGSGGAGAT	1800
Db	1746	GGTACTGAACCAATCAATTTAGAAAGACTTCAATACATCCCGGTATGSGGAGAT	1805
QY	1801	CTAACTGAGAAGGCTGAGAAAGTTTCTATATCTCCATPAPTAAAGACAAGGTGAC	1860
Db	1806	CTAACTGAGAAGGCTGAGAAAGTTTCTATATCTCCATPAPTAAAGACAAGGTGAC	1865
QY	1861	ACACCAGAAACCCGAAGCATCTTCCAAATCTGGCTATTCTGATTAATCTATTGAGTACTT	1920
Db	1866	ACACCAGAAACCCGAAGCATCTTCCAAATCTGGCTATTCTGATTAATCTATTGAGTACTT	1925
QY	1921	CAATCAGTTTAGATTTGATGGAATTGAAATTAATCTTAATGTTGGGGTTAAAGCATCA	1980
Db	1926	CAATCAGTTTAGATTTGATGGAATTGAAATTAATCTTAATGTTGGGGTTAAAGCATCA	1985
QY	1981	AGCTTCTTAAACTCAATAGACCAAAATCCAAACGGCTAGAAATGATTGCTGCACACATCA	2040
Db	1986	AGCTTCTTAAACTCAATAGACCAAAATCCAAACGGCTAGAAATGATTGCTGCACACATCA	2045
QY	2041	TACTTAAAGTCACAAATTTGGATTAGCTAGAACATCTGATTTACCAACCAACCAACATTC	2100
Db	2046	TACTTAAAGTCACAAATTTGGATTAGCTAGAACATCTGATTTACCAACCAACCAACATTC	2105
QY	2101	GGAAACAATCAACCAAGTTATTTCACTATCACCTGGTGATCAGTTCTCATCAATTPAGAT	2160
Db	2106	GGAAACAATCAACCAAGTTATTTCACTATCACCTGGTGATCAGTTCTCATCAATTPAGAT	2165
QY	2161	ATTAGAACAAATCTCCCTGCTTACACAGTTATGATGTAATCTTATATTCACAAATGA	2220
Db	2166	ATTAGAACAAATCTCCCTGCTTACACAGTTATGATGTAATCTTATATTCACAAATGA	2225
QY	2221	AAATCTAGTGTATTATCAATTAAGATTACTGATCAATCAAGTAAACCGTGAGCTCAAGCTCA	2280

```

|||||
2226 AATCTAGGTTTATACATTAAGATAGCTGACCAAGTAACCTGATGGCTTAAGCTCA 2285
2281 TTCAGTCCAAAGTTTATAGAGTTTAAATGTTGATCTTACCTTTATTAGAC 2340
2286 TTCAGTCCAAAGTTTATAGAGTTTAAATGTTGATCTTACCTTTATTAGAC 2345
2341 AATCTATCTATACAGTAATAGCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2400
2346 AATCTATCTATACAGTAATAGCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2405
2401 TCTCTGATCATATCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2460
2406 TCTCTGATCATATCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2465
2461 GAAGTCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2520
2466 GAAGTCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2525
2521 GAGTACTTATTTCAAAATGGTTCATAGTCAAGTGTGAGAACTGCTTACAAACCA 2580
2526 GAGTACTTATTTCAAAATGGTTCATAGTCAAGTGTGAGAACTGCTTACAAACCA 2585
2581 AGCTCTTAAACAGTTTATGATGCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2640
2586 AGCTCTTAAACAGTTTATGATGCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2645
2641 GTGCTGATCTGTAATGTTGATCTTACCTTTATTAGAC 2700
2646 GTGCTGATCTGTAATGTTGATCTTACCTTTATTAGAC 2705
2701 TACTTAGATGTAATGTTGATCTTACCTTTATTAGAC 2760
2706 TACTTAGATGTAATGTTGATCTTACCTTTATTAGAC 2765
2761 TTTAACAATTTGGGCTTACCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2820
2766 TTTAACAATTTGGGCTTACCTGCTGTAATGTTGATCTTACCTTTATTAGAC 2825
2821 GGTATCTGTAATGTTGATCTTACCTTTATTAGAC 2880
2826 GGTATCTGTAATGTTGATCTTACCTTTATTAGAC 2885
2881 AGAATTTCAAGACAAAGGTTCAAAACATTTCAAAAGTTGATGATCTTACCTTT 2940
2886 AGAATTTCAAGACAAAGGTTCAAAACATTTCAAAAGTTGATGATCTTACCTTT 2945
2941 GCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3000
2946 GCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3005
3001 GCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3060
3006 GCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3065
3061 GCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3120
3066 GCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3125
3121 TCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3180
3126 TCTGTTGTTCAATTTACAAAGATTTTACCCAACTGCTTACCTTTATTAGAC 3185
3181 AAGAAATTA 3189
3186 AAGAAATTA 3194

```

RESULT 6
 AX65242 3129 bp DNA linear PAT 26-MAR-2003
 LOCUS
 DEFINITION Sequence 79 from Patent EP1275716.
 ACCESSION AX65242

```

VERSION AX65242.1 GI:29290367
KEYWORDS Mycoplasma gallisepticum
SOURCE Mycoplasma gallisepticum
ORGANISM Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
REFERENCE 1
AUTHORS Okuda, T., Saito, S., Dorsey, K.M. and Tsuzaki, Y.
TITLE Modified dna molecule, recombinant containing the same thing, and
JOURNAL uses thereof
patent: EP 1275716-A 79 15-JAN-2003;
Zeon Corporation (JP)
FEATURES
source 1..3129
/organism="Mycoplasma gallisepticum"
/mol_type="genomic DNA"
/db_xref="taxon:2096"
/note="Modified mgc3 gene"
BASE COUNT 1078 a 581 c 541 g 929 t
ORIGIN
Query Match 95.9%; Score 3057.8; DB 6; Length 3129;
Best Local Similarity 98.7%; Pred. No. 0; Mismatches 42; Indels 0; Gaps 0;
Matches 3083; Conservative
65 TTGGTCTGCAACGCTTGGCTTTAAGCAATCAGATTAAGTAAGCAATTAAGCAATTAAG 124
5 TCGGCTTCAACCTTTGGCTTTAAGCAATCAGATTAAGTAAGCAATTAAGCAATTAAG 64
125 TTAATCAAGCAAGCAAGCAATGATGATTAATTTCTTTAGACTGCAAGTCTGGCAAAATG 184
65 TTAATCAAGCAAGCAAGCAATGATGATTAATTTCTTTAGACTGCAAGTCTGGCAAAATG 124
185 GTTGTGTTCAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 244
125 GTTGTGTTCAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 184
245 GAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 304
185 GAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 244
305 ATTGTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 364
245 ATTGTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 304
305 ATTGTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 364
365 TTGATCAAGCAAGCAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 424
245 TTGATCAAGCAAGCAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 304
305 TTGATCAAGCAAGCAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 364
425 ATTTAATAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 484
365 ATTTAATAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 424
485 TTGTAATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 544
425 TTGTAATTTGTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 484
545 AATTTAATTTTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCA 604
485 AATTTAATTTTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCA 544
485 AATTTAATTTTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCAAACTCA 544
605 CAATCTTACCTTAAGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 664
545 CAATCTTACCTTAAGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 604
665 TACCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 724
605 TACCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 664
725 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 784
665 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 724
785 CTCAAACTTAATAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 844

```


|||||
Db 725 CTCMAACTAAACCCACTACTGATGATTAATCAAAATCTTCAACTTTTAAATTCAGAGCAATCG 784
845 CTGGTCGCAAAACATAGATGATTCATATTTGAATGCAAGCATAGATTAATTAACATCTT 904
785 CTGGTCGCAAAACATAGATGATTCATATTTGAATGCAAGCATAGATTAATTAACATCTT 844
905 TCCAAATAGATGAAAAATTTGGTTATTCAGAAATGAGCTGGTCTGGAAGAAATAAAAAT 964
845 TCCAAATAGATGAAAAATTTGGTTATTCAGAAATGAGCTGGTCTGGAAGAAATAAAAAT 904
965 TTACAAGATTAGCTACTGGAAGTTGCCAAGCAACGAAGATTTGATTCCTTGACATAC 964
905 TTACAAGATTAGCTACTGGAAGTTGCCAAGCAACGAAGATTTGATTCCTTGACATAC 964
1025 CCGGAGCTCCACAGATTACTTTTAAAGAGATTCATTAACGATTTTCAAGCTTACT 1084
965 CCGGAGCTCCACAGATTACTTTTAAAGAGATTCATTAACGATTTTCAAGCTTACT 1024
1085 TAACTCAGTTAATTTCTTATATCATTCATTTGGTATGATTAATTTTGGTACCTCG 1144
1025 TAACTCAGTTAATTTCTTATATCATTCATTTGGTATGATTAATTTTGGTACCTCG 1084
1145 AATTACCATCATTTATGCTACTATTCATTCCTCAACCTGATTTATCTGATCTTGA 1204
1085 AATTACCATCATTTATGCTACTATTCATTCCTCAACCTGATTTATCTGATCTTGA 1144
1205 ATCAAGTTAAACAGATGATTTGAAGCTTCAAGCTGATTAAGCTACCAACCAACAG 1264
1145 ATCAAGTTAAACAGATGATTTGAAGCTTCAAGCTGATTAAGCTACCAACCAACAG 1204
1265 GAACAAGACACAGCTGATACATCTAGTGTTCACACAGTGTGGAACAGAAATCTA 1324
1205 GAACAAGACACACAGCTGATACATCTAGTGTTCACACAGTGTGGAACAGAAATCTA 1264
1325 CTACACTCTCTCAACAGTTTCTAATCTACTTAAATCTAATCTAATCTAATCTAATCT 1384
1265 CTACACTCTCTCAACAGTTTCTAATCTACTTAAATCTAATCTAATCTAATCTAATCT 1324
1385 ATATGATTAACCACTTCTGCAAAACAAATGATGAATTAATGGAAGATCTTAACGTTA 1444
1325 ATATGATTAACCACTTCTGCAAAACAAATGATGAATTAATGGAAGATCTTAACGTTA 1384
1445 TTGAACAGAAATATATGCTGATACAGATTAGTATTCAAAATGAAATCCAAATCTA 1504
1385 TTGAACAGAAATATATGCTGATACAGATTAGTATTCAAAATGAAATCCAAATCTA 1444
1505 ATGCAAGAACTTTATCCGAACACAAATGAGTGTGTTGTTTACTCAACAGTTCAA 1564
1445 ATGCAAGAACTTTATCCGAACACAAATGAGTGTGTTGTTTACTCAACAGTTCAA 1504
1565 GAGTAGTTTAAAGAGCTTCTTAAAGGAGATCAAGCTGCAAGTGAACCTTCCAACTT 1624
1505 GAGTAGTTTAAAGAGCTTCTTAAAGGAGATCAAGCTGCAAGTGAACCTTCCAACTT 1564
1625 TCTTATACGATTTGGTTATTTAGATACCAACAACTAGAACAGAACTTCTGTAGC 1684
1565 TCTTATACGATTTGGTTATTTAGATACCAACAACTAGAACAGAACTTCTGTAGC 1624
1685 GAACATATAGCTTTTAAACACAGCCCTTACAGCTATTTAGATTTCCAAAGTAGTA 1744
1625 GAACATATAGCTTTTAAACACAGCCCTTACAGCTATTTAGATTTCCAAAGTAGTA 1684
1745 CTGAAGCAATTAAGCTTTTAAAGAGCTTCAATTAACCTGTTATGAGGAGTAGTA 1804
1685 CTGAAGCAATTAAGCTTTTAAAGAGCTTCAATTAACCTGTTATGAGGAGTAGTA 1744
1805 CTGAAGCAATTAAGCTTTTAAAGAGCTTCAATTAACCTGTTATGAGGAGTAGTA 1804
1745 CTGAAGCAATTAAGCTTTTAAAGAGCTTCAATTAACCTGTTATGAGGAGTAGTA 1804
1865 CAGAAAGCGAAGCATCTTCAATCTGCTATTTCTGATTAATCTATGACTACATTCAT 1924
|||||

Db 1805 CAGAAAGCGAAGCATCTTCAATCTGCTATTTCTGATTAATCTATGACTACATTCAT 1864
1925 CAGTTTATGATTTGATGATTAAGATTAATTAATTAATGTTGGGTTAAAGCATCAAGCT 1984
1865 CAGTTTATGATTTGATGATTAAGATTAATTAATTAATGTTGGGTTAAAGCATCAAGCT 1924
1985 TCTTAACTCAAAATAGACCAAAATCCAAAGGCTTACAAATGATTTGCTGCAACATCTACT 2044
1925 TCTTAACTCAAAATAGACCAAAATCCAAAGGCTTACAAATGATTTGCTGCAACATCTACT 1984
2045 TTAAGATCACAATTTGATTAAGCTAGAACATCTGATTTACCAACCAACCACTTTGGAA 2104
1985 TTAAGATCACAATTTGATTAAGCTAGAACATCTGATTTACCAACCAACCACTTTGGAA 2044
2105 CAACCTCACAAGTTATTTAGATTAAGCTAGAACATCTGATTTACCAACCAACCACTTTGGAA 2164
2045 CAACCTCACAAGTTATTTAGATTAAGCTAGAACATCTGATTTACCAACCAACCACTTTGGAA 2104
2165 GAACAATCTTCCCTGGTAAACCAAGTTATGATCTTCTTATTCACAAATGAAATTAATTAAT 2224
2105 GAACAATCTTCCCTGGTAAACCAAGTTATGATCTTCTTATTCACAAATGAAATTAATTAAT 2164
2225 CTAGTGTATTAATTAAGATTAAGCTAGAACATCTGATTTACCAACCAACCACTTTGGAA 2284
2165 CTAGTGTATTAATTAAGATTAAGCTAGAACATCTGATTTACCAACCAACCACTTTGGAA 2224
2285 GTCCAACAAGTTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2344
2225 GTCCAACAAGTTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2284
2345 CATTCTATACAGTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2404
2285 CATTCTATACAGTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2344
2405 CTGATCATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 2464
2345 CTGATCATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 2404
2465 GTTCTGTGTATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2524
2405 GTTCTGTGTATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2464
2525 ACTTAATTCAAAATGAGGCTTCAATGATCAAGTGTGCTGTAACCTTCTGTTAAACCAACGCT 2584
2465 ACTTAATTCAAAATGAGGCTTCAATGATCAAGTGTGCTGTAACCTTCTGTTAAACCAACGCT 2524
2585 TCTTAAACAGTTAGTTAGTACTCAGTCTGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2644
2525 TCTTAAACAGTTAGTTAGTACTCAGTCTGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2584
2645 TTGATCCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2704
2585 TTGATCCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2644
2705 TAGATGTAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 2764
2645 TAGATGTAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 2704
2765 ACAACTTTGGGCTTACCTTCAATGAGTGTGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2824
2705 ACAACTTTGGGCTTACCTTCAATGAGTGTGCTGATGCTGATGCTGATGCTGATGCTGATGCT 2764
2825 TTCTGCAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2884
2765 TTCTGCAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2824
2885 AATTCAAGCAAGGCTTCAAAACCAATTCACAAAGTTGATTAACCTTGCATGCTGCTG 2944
2825 AATTCAAGCAAGGCTTCAAAACCAATTCACAAAGTTGATTAACCTTGCATGCTGCTGCTG 2884
2945 TTGCTGCAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 3004
2885 TTGCTGCAATTAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTA 2944

Db 6867 AAAATATACAGATAGCTAGCTGGAAGTTTGCCAAACAGAGATATGAATCTTG 6926
QY 1019 ACATACCCGGGACTCCACAAGTTACTTTAAAGAAATTCAGTTAACTTTTCAAGAC 1078
Db 6927 ACATACAGGAGCTCCAGAAGTTACTTTAAAGAAATTCAGTTAACTTTTCAAGAC 6986
QY 1079 TATACCTAATCAGTAACTTATCTTATCATTCATTCGTGATAGTATTTATTTTGGTA 1138
Db 6987 TATACCTAATCAGTAACTTATCTTATCATTCATTCGTGATAGTATTTATTTTGGTA 7046
QY 1139 CCTGTAATTAACCATTAATGATGATATTCATTCATTCATTCATTCATTCATTCACCG 1198
Db 7047 CTCTGTAATTAACCATTAATGATGATATTCATTCATTCATTCATTCATTCATTCACCG 7106
QY 1199 CTCTGTAATTAACCATTAATGATGATATTCATTCATTCATTCATTCATTCATTCACCG 1234
Db 7107 CTCTGTAATTAACCATTAATGATGATATTCATTCATTCATTCATTCATTCATTCACCG 7141

RESULT 8

AB033210

LOCUS AB033210 1131 bp DNA linear BCT 07-APR-2000
DEFINITION Mycoplasma gallisepticum gene for 120-kDa membrane protein MG3,
partial cds, strain: S6.

ACCESSION

AB033210

VERSION

AB033210.1

GI:7527370

KEYWORDS

120-kDa membrane protein MG3.

SOURCE

Mycoplasma gallisepticum

ORGANISM

Bacteria: Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

REFERENCE

1 (bases 1 to 1131)

AUTHORS

Yoshida, S.

TITLE

Mycoplasma gallisepticum S6-strain gene encoding a 120-kDa membrane protein

JOURNAL

Published Only in Database (2000)

REFERENCE

2 (bases 1 to 1131)

AUTHORS

Yoshida, S.

TITLE

Direct Submission

JOURNAL

Submitted (04-OCT-1999) Shigeto Yoshida, Jichi Medical School, Department of Medical Zoology, Yakushiji 3311-1, Minamikawachimachi, Tochigi 329-0498, Japan (E-mail: shigeto@jichi.ac.jp, Tel: 81-285-58-7339, Fax: 81-285-44-6489)

FEATURES

Location/Qualifiers

1..1131

source

/organism="Mycoplasma gallisepticum"

/mol_type="genomic DNA"

/strain="S6"

/db_xref="taxon:2096"

1..1131

/gene="mgc3"

1..>1131

/gene="mgc3"

/codon_start=-1

/transl_table=-4

/product="120-kDa membrane protein MG3"

/protein_id="BA94377.1"

/db_xref="GI:7527371"

/translation="MNISKKLKSLVLLIGLAVFGLTGLSAPFGKQSDKSDNTQVNO
ARTLDANSVRLAGLGGSLFNTVLDVDNFTTANGTILKIDSTPKPYGLDSD
FAGYKQIVSDITTSRNRFQROTARYALLVNDANVHLKRNINSNIGRNNS
KFVIGVDNPAHYIRFTDGTKEFTKOTGEIVNDPILADPILPKDLHDWVNLVIO
RKLIPDVNTAVPMPVGRVSGNSATDGPEDGNGMPSANDPIATSKSGSONPSSFN
SSAOPCTNNRKYSEVNVKHKIKTSFOLDKFTYPTENTGSEKNITRLATGSLPNNEG
YWLIDLPGEPEVLTKEDSVNFSRLYLNVSLSFTGDSLYIF"

BASE COUNT

411 a

169 c

197 g

354 t

ORIGIN

Query Match

31.1%

Score 993.2

DB 1

Length 1131

Best Local Similarity 92.9%

Pred. No. 1.7e-157

Matches 1053

Conservative 0

Mismatches 78

Indels 3

Gaps 1

QY 1 ATGAATATTCTTAACCACTTAAGATTATACATGATAGTGATGATGATTTGCA 60
Db 1 ATGAATATTCTTAACCACTTAAGATTATACATGATAGTGATGATGATTTGCA 60
QY 61 GCTCTGTTGCTGCAAGCTTTGGCTTTAAGCAATCAGATTAAGATTAACGACCA 120
Db 61 ACTCTGTTGCTGCAAGCTTTGGCTTTAAGCAATCAGATTAAGATTAACGACCA 120
QY 121 TTAGTAAATCAACCAAGCAAGCTAGATGCTAATTCGTGTAGACTGAGCTTTGACAA 180
Db 121 TTAGTAAATCAACCAAGCAAGCTAGATGCTAATTCGTGTAGACTGAGCTTTGACAA 180
QY 181 AATGTTGCTGTTCAATCAGTCTTAGAGATGTTGATGATTAACCTTAACGACGCT 240
Db 181 AATGTTGCTGTTCAATCAGTCTTAGAGATGTTGATGATTAACCTTAACGACGCT 240
QY 241 AATGAAACAATTAATCAATAGATGTTTACTTAACCAATTAATGTTTAACTTAAGT 300
Db 241 AATGAAACAATTAATCAATAGATGTTTACTTAACCAATTAATGTTTAACTTAAGT 300
QY 301 GATGATTTGCTGATACCAAGTAAACCAATAGTTTCAATTAACCAATTAACCAAGT 360
Db 301 GATGATTTGCTGATACCAAGTAAACCAATAGTTTCAATTAACCAATTAACCAAGT 360
QY 361 AGATTTGATCAACCAAGCAAGATATTAATGCTGCTGTTGATGATGATGATGATGAT 420
Db 361 AGATTTGATCAACCAAGCAAGATATTAATGCTGCTGTTGATGATGATGATGATGAT 420
QY 421 GTTCATTTTAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
Db 421 GTTCATTTTAAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
QY 481 AAGTTTGTAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 AAGTTTGTAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 541 ACTAAATTTAATTTTCAACCAAGCAAGTAAATTTTATGATGATGATGATGATGATGAT 600
Db 541 ACTAAATTTAATTTTCAACCAAGCAAGTAAATTTTATGATGATGATGATGATGATGAT 600
QY 601 GCGCCAACTTACCTTAAGATTTTACACCCAGATTTGATTAATTAATTAATTAATTAAT 660
Db 601 GCGCCAACTTACCTTAAGATTTTACACCCAGATTTGATTAATTAATTAATTAATTAAT 660
QY 661 ATCTTACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
Db 661 ATCTTACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
QY 721 ACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
Db 721 ACAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
QY 781 AATAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Db 781 AATAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
QY 841 ATAGCAACGCTTAAGACCGGCTTCAAAAT---AACCCCAAGTTCAATTCAGCGCT 897
Db 841 ATAGCAACGCTTAAGACCGGCTTCAAAAT---AACCCCAAGTTCAATTCAGCGCT 897
QY 897 TCTTTTCAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
Db 897 TCTTTTCAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
QY 961 TCTTTTCAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
Db 961 TCTTTTCAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
QY 1021 ATACCGGAGCTGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080
Db 1021 ATACCGGAGCTGCAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1080
QY 1081 TACTTAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1134
Db 1081 TACTTAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1134

TITLE
JOURNAL
COMMENT
FEATURES
source

Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Saudak,D.M.,
Phillips,C.A., Merrick,J.M., Tomb,J.-F., Dougherty,B.A., Bott,K.F.,
Hu,P.C., Lucier,T.S., Peterson,S.N., Smith,H.O. and Venter,J.C.
Direct Submission
Submitted (19-Oct-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Nov 5, 1998 this sequence version replaced gi:1045891.
Location/Qualifiers
1.15787
/organism="Mycoplasma genitalium"
/mol_type="genomic DNA"
/isolate="G37"
/db_xref="taxon:2097"
1264.2016
/gene="MG186"
1264.2016
/gene="MG186"
1264.2016
/note="similar to GB:U00089 SP:P75265 PID:1673668 percent
identity: 44.35; identified by sequence similarity;
putative"
1.15787
/codon_start=-1
/transl_table=4
/product="lipoprotein, putative"
/protein_id="AAC71405.1"
/db_xref="GI:3844783"
/translation="MKGLKPNFSLGALFLTLSPATACIAKPKYNNRPNSEGLAR
LKRVRNHRDGTLEYSFANHOKPRLIYALIDPEKAVLSQRSELEKANKATE
FAKLPIGSEWMTPLNSYSDREVAAYEFKPNQLHESFAVEMVANGALDIAG
NDFVFSDDPPRPLKIVGIELANGNANFRKKIFSYLNSFOSITMYVQGVGD
QSWTRYLAPSNDESKTGLGILTYELKLNNG"
2009.3766
/gene="MG187"
2009.3766
/gene="MG187"
/note="similar to GB:U00089 SP:P75264 PID:1673667 percent
identity: 61.74; identified by sequence similarity;
putative"
1.15787
/codon_start=-1
/transl_table=4
/product="ABC transporter, ATP-binding protein"
/protein_id="AAC71406.1"
/db_xref="GI:3844784"
/translation="MANKSAIELEKNIYVDFGESVAIDNINLSEYKHQVLSLIPSGC
GKTTTLAVAGLIKPTSGOVLENGYDVKRPPORRKLGLVONVALPNNHVENIVF
PLYDSMKQAVLEKNSVANHINCLITNGASVOEIDOLNLEHSEIKRKLQKOT
INDLANSYFKNLELNFANLKLIPSKHOPALINIKKOTLQINLEAEFKRKLQKOT
PIKSGVEHNAKLQAIKOHSEYKORLKHRYKTVELKOTLVENLKLTKKASISQETL
IKOSSDYTKLQKRLKIVPEPQKRYKVPUNOLIKRYSIKTQKIDLOLNEIOIK
TRIVSIKOPINKTALAVANKLAITKRLRPDKISGGQOQAVAKARAVRPKILLMD
EPLSNIDAKLRVOTROWIROFOEQLITVTVHDEEAMSIISVYQMSIGKRYQOIG
TPSEIKLPAHEVAREFLGTPPEMNIIIECSVKNNQLENNHLLTSESEKLEKLVGF
RYQOLVYTNKSSLOKININENLCKHVAITSLDITLSMKLELHSLKVGDSLNI
IKANNLHFPDIDKORIEI"

gene
CDS

4682.5686
/gene="MG189"
/note="similar to GB:U00089 SP:P75262 PID:1673665 percent
identity: 69.67; identified by sequence similarity;
putative"
1.15787
/codon_start=-1
/transl_table=4
/product="ABC transporter, permease protein"
/protein_id="AAC71407.1"
/db_xref="GI:3844785"
/translation="MFKMLKHNHOPHSLOGLDPLDFWKPFLFLPALTLITFL
ILPPLSLQGEFANSDLUDLSOSFSLRPFODLFSBNFVLGLRNSFLYSLISLPS
ITLAIIVASAVFYKRLKLGQVTFELPYTSVAISIAFYIIDSASGLDNTVN
VNTKMLDSGSDTYNNALMALIITGCVKMLAFNVLIISTAMLSVNPOLYVASLDSNP
VROFKRILSPRLPLFTLTLILGSMQVPLALPENKREBAVANGANSILYIFQO
IOSGNTNLAGAATLVLFVGLVGLVLRNGFYILEMLQMKIKOLYVQKOLITY"
4682.5686
/gene="MG189"

CDS

4682.5686
/gene="MG189"
/note="similar to GB:U00089 SP:P75262 PID:1673665 percent
identity: 69.67; identified by sequence similarity;
putative"
1.15787
/codon_start=-1
/transl_table=4
/product="ABC transporter, permease protein"
/protein_id="AAC71408.1"
/db_xref="GI:3844786"
/translation="MARTLNGCSEKLSFMRKNNLRFTSWINQKRFYQDLSLKTRS
IKQIVLVETLVGLVFGILVIVPPPLMVVYFASDERALDTRPILMPDSWDFNF
SKVLDGKRYNAIVNVLVYSLTLFTTGMGFSLRKKRYKLVWFFFLYLI
LPSGALIGORYRIVANMNNPSLIVGLDILPVSVCSPGMRSEALPSQLE
SALIDGCGFNYFLKIALPMVYKSTWVGLIWFSAANSNLPMLILGKRVLDNIMM
VLOQSIDANSDEQIRTLNLKMSAAILALPLWFIIFYLFHRIMALKNNANTIKG
"
5726.6811
/gene="MG190"
5726.6811
/gene="MG190"
/note="similar to PID:1196657 GB:U00089 SP:P75144
PID:1673660 percent identity: 60.38; identified by
sequence similarity; putative"
1.15787
/codon_start=-1
/transl_table=4
/product="mgpa protein (mgpa)"
/protein_id="AAC71409.1"
/db_xref="GI:3844787"
/translation="MGKSTIENHCVINDYLISVCHLMHCASDHYHOMRGFTFIYV
KSGITEAINALIKOPDKIVIFPHVRPDGCLDAQGLHLLKANEKKVEYCVGNND
LPSFIMTFTNOIDSEFLKEALAIYVDNAVYKRIIRLRLKRLKALRIDHPND
DLWTSFNPEFSEVACCEQIVEMATYAKWTIPVATLLYIGIYDSNRFLYSSTR
TLVLAAILYAKARDIRIVHDHINHTSLADLKFKKYVNHFTQGOVIFYICTKIQK
LRMTAQCARVNLNSIADYKIMLPTEIQANNEIRLDRSGINVRDIAIKYGGGHN
NASGAIITNKQISDVSDCVKRIYV"
6818.11152
/gene="MG191"
6818.11152
/gene="MG191"
/note="similar to SP:P20796 GB:M1421 GB:U01694 GB:U01779
GB:U02159 percent identity: 98.89; identified by sequence
similarity; putative"
1.15787
/codon_start=-1
/transl_table=4
/product="adhesin p1 (mgpa)"
/protein_id="AAC71410.1"
/db_xref="GI:3844788"
/translation="MHQKRLAKRSMAFLAALLTGLVITGVGVFLPNOKRORSYS
NFAVQPOLSKYHOQAVNDENTLTPTMNNNNFSLKLTGENPDSFGVLRQNDNLIS
VKNSSDNLKLYLNAVEKYLDCQONFPIRRYDNNGRALYDINLAKMENSIVQGLNG
EPIIDPFEGFGLTGNAPDMMELRGKPYEVOSHPNLYFVLVAPKALAEYHNAN
QYKESLEVKATQSSFNTPQRLQKDSYKQSGKSEKLSFTTASSMSGKATSTRAKA
LKEVERGSSQSDSLKNDFAKPLKHNKSSQEVAKLEKREFTAMPHLITTDQIAEK
MGATVVSFYDPAESENHTAFGLVDHIDPKVMVNPYPSKTRKNNHGHGWONANL
ILQTFGFNPRHPEMPEDEGQAKADNTSPGKVDYDTHKDGKRNSSSPILAPRAY
FANIGMVAIGNSVFIPEGNGHATKMTFNLISIGVRIRYTDNFEKSSVGTGMPAVL
KQGLEHPTQNGKLDPLIGNHMEVEVPEPMDIPVGMALDDDEGSKRGHYLELN
FDDLEKHLNLVAQOGGILREDIOTFPGVGNPDPIDVGMALDDDEGSKRGHYLELN
NDPIDQNVNNDVEALISYKNTDRLKRYVYRSGIYAMOLFMSKRLNTPLSANF
VENSAYAPNSLFAALINEDLTLGLSDKLFYKKNMPEANEDRNLQSLNPPTFSN
ARYLVVQRFETTGPNLDSSTFQDLPIDPLWNGKQVQSANQKNGIPEQVRSNNS
INVGVSMITQHLNKNENTRWVPIPNFSDIWMGATRVQSANQKNGIPEQVRSNNS
TPDPNSDNKPYSGSSSKPTTYALPNLSISPDSMDINALFTFKNNRPNOLITGFE
LATTIYVLINKSGDSNDQNKSEOKMKDTENEGCNLFGVEVGLNVAALHLITGFE
GLTNWSTDREKIGKADSSSSSTLVSGLNMTSDQVGNLVAVINDTSFGQDGMFT
PTDFIRIRGTGIDITLSSLODOTIIMADOPMTSFGYLDSPKPSIMDPALAKSLP
NSSYTDIRNGYDITLSPOLQPNKRYAYQTTNTYKILBEVDAVTSATIMDLKLT
KNIKAKIGKSTASQGNNGGVSQITNTITTTNGISGKEEBSIOAETLKFFDSK
ONKSEIGDSDSTFKMDKLTGVSFTPLVNLINQOGATSDSPKTSERKQOIDEF
RLTLPTELEFNTMEVYDQVPLVNLPSGFDASIKLAVTISVNOPLGVRLEP
KDPQTOGFIIVLNAASSTPGQIVOPENQADAVPLVIVTPVIVILISVTLGRTIGIP
MHRKRALQAGFPLSKKKVDVLKAVGSFKEKILNHTGISNAPKRLKQATPTKPKTI

gene
CDS

```

/codon_start=1
/translation_table=4
/product="mgpc protein (mgpc)"
/protein_id="AAC71411.1"
/db_xref="GI:3844783"
/translation="MTMTKRLQIKKAYWLLDLPFLALANTFLYEDSKNRYAATPPA
TPIDRSKSDVLSLAQDSVLAQDQIHTNINFLYFKSRDYKVEKSSNNISFDSST
SOGKPEYVEYFNINSGIKRMWYQYQDLPVNPSPNNQVKKLILEQFLPKYKLS
SLIAKEKQIQRVEYHIGSGAQNMTSGQRHDLNNPSPNNSTGKRLTGNAAYKRIE
SMYLPYPIIDTQKQKGRDSSGMSSTENAKNDASVYSGGSSSTPKYLTATQALIE
SGLLEPNDQTPRNVITQLIYASTSKAVLANNIVMGNSFDSMTYVWESDVAWES
KKPFWENQDQKQKQPVENQIGYKETTISINSHNSKSFQPPYLLISGIDSVNW

```

QY	2770	TTTGGGCGCCTTACCTTCATGGGTAGTGCCTACACACAATTTGGTAGTACATTCTT	2829
Db	13950	TTTGCACACACATCCCTCGCATGGGTGATCCCTGATACAGAGAGTCTTCAGTTGGGATCTTG	14000
QY	2830	GCAATTATGATCMATCTAGGATTAGCTATAGTGGGTATTCCTTTAAGAGCTCAAGAAGAAATTA	2889
Db	14010	TTTATCTCTTGTAGTCTTAGGACTTGGATTTGGATCCCAATGTACAGGGTAGGAATCTC	14066
QY	2890	CAGACAAAGGGTTCAAAACAACATCTCAAAAACCTTGATATACCTTGCTGCTTGGT	2949
Db	14070	CAAGATGCATCGTTGTTGTAATGTCTTTAAAAAGTTGATACACTCACAACTGCTGCTGGT	14121
QY	2950	TCAGTTTACAAAGATTTATTTACCCAATGCTCTAACGTTAAGAAAACCTGCTGCTTTA	3009
Db	14130	AGGTGTACAAAAAGATTATTATCCCAACTGCTGTGGAAAAAAGCACTAGTGCATTTG	14181
QY	3010	GGGTCTGGTAAATCTGGTGATAGAAAACCTGCTGCTGCTAAACCTGCTGCTCAC	3068
Db	14190	AAAGCTCTATACCTTAAGTGTAAAAAAACCTGCTCTTTTAAAAACCACTGCTTCAACC	14248

OY		2770	TTTGGGCGCTTACCTTCATGGGTAGTGCCCTCAGACAATTGGTAGACATPPTAGTAATTCYT	2822
Db		28703	TTTGCAACACTCCCCTGCATGGGTATGCCCTGTATCATAGTAGTCTTTCAGTTGGGATCTTG	2876
OY		2830	GCAATTATGATCATCTTAGGATTTGTATGGSTATTCCCTTTAAAGCGTCCAAGAATAA	2883
Db		28763	TTTTATCTTGTAGCTTAGGACTTGGGATTTGGATCCCAATGATGAAGGCTAGAAAATCTC	2888
OY		2890	CAGACAAAGGTTCAAACACAACTCACAAAAAGTTGATACCCTTGACTGCTGCTTGGT	2943

	RESULT 13
MYCATTP	
LOCUS	
DEFINITION	9691 bp DNA linear BCT 15-FEB-1996
ACCESSION	M.pneumoniae p1 operon encoding p1 surface protein, ORF-4, and
VERSION	ORF-6 proteins, complete cds.
KEYWORDS	M21519 M20916
SOURCE	M21519.1 GI:150138
ORGANISM	p1 attachment protein: p1 surface protein.
REFERENCE	Mycoplasma pneumoniae
AUTHORS	Mycoplasma pneumoniae
	Bacteria: Firmicutes: Mollicutes: Mycoplasmataceae; Mycoplasma.
	1 (bases 1054 to 5937)
	Itanahne, J.M., Denny, T.P., Loeschel, S., Schaper, U., Huang, C.H.,

TITLE	Botl, K.F. and Hu, P.C. Nucleotide sequence of the P1 attachment-protein gene of <i>Mycoplasma pneumoniae</i>
JOURNAL	Gene 64 (2), 217-229 (1988)
MEDLINE	88297153
PUBMED	2841195
REFERENCE	2 (bases 1 to 9691)
AUTHORS	Inamine, J.M., Loechel, S. and Hu, P.C.
TITLE	Analysis of the nucleotide sequence of the P1 operon of <i>Mycoplasma pneumoniae</i>
JOURNAL	Gene 73 (1), 175-183 (1988)
MEDLINE	89211947
PUBMED	2468577
COMMENT	Original source text: <i>Mycoplasma pneumoniae</i> (strain M129) (tissue

FEATURES
source Location/Qualifiers
1. 9691
Library: ATCC 29342) DNA.
Draft entry and computer-readable sequence for [2] kindly provided
by S. Loechel, 16-NOV-1988
Location/Qualifiers

```

/organism="Mycoplasma pneumoniae"
/mol_type="genomic DNA"
/strain="M129"
/db_xref="taxon:2104"
/tissue_lib="ATCC 29342"
54..9616
mrna
/product="P1 operon mRNA"
316..1041
CDS

```

```

/note="ORF-4; putative"
/codon_start=1
/transl_table=4
/product="unknown protein"
/protein_id="AA08324.1"
/db_xref="GI:1196657"
/translation="MQQALAVDYANKERIECDLLDQNOFKAVLRIDHNPEDDILN
TTHNFVPSAYIAAEQYVDLAVQAKMLSPPAALVIGITDSENREYVSTTSRTLYE
IGSLVYRAQANIKAIHDELNHTSLKDLPQYVRKNRQFQNTVYFAADKFOKKLVK
TPLECAVNTLANTEQPHIMFLFEELCKNHYRVFRRSNGITNREVALKYGCGHIQNS
GAYLAKSRDIIIRVQDCQKDIADV"
1054 . . 5937

```

```

/codon_start=1
/transl_table=4
/product="P1 Protein"
/protein_id="AA08325.1"
/db_xref="GI:150139"
/translation="MHQTKRTALSKTWIILITATPASIANGLTIVVGHFTSTTTLLKRO
FESTRPDEVALRHTNAINPLIPWYRNNTFSSLPITGEPGAMALVRDASAGITPA
GSGSGOQTYTPDTPREALITASTPFLALRYDLAAGALIDFSKINPDTPDQGTITL
ENPPGGTGLSGAAPQOWNNEKKNKPYEVAODPSNPYFAFVILVRSYVYEQDLRGGLG
LPQRTSGGSLSTTGAMFGLKYNVNAADVAKSEKIQGAEATGSSSTTSGGSGSTGRK

```

[illegible]

Query Match	3.2%	Score 102.2;	DB 1;	Length 9691;
Best Local Similarity	55.58;	Pred. No. 1.2e-07;		
Matches 197; Conservative	0;	Mismatches 150;	Indels 0;	Gaps 0;

QY	2737	ACACATTAATTAGTAAACATCTCTTTAATAACAACGTTGGGCGCTTACTTCACATGGGTACTG	2196
Db	9234	AACTTGATTTCCCTTCCTCCGATTTTTCGTGGCTTGCTCTTTACCGTCGCGGTCAAT	9293
QY	2197	CCACAGCATTTGGTAGTACATTAGTATTTCTTGCATTAATGATCATCTTAGGATTAGCT	2856
Db	9294	CCGGTATCACTCGCTTCAATCGGTGGCATTTCTTAATCCGTCACTTCTTAAGCCCTTGGT	9353
QY	2857	ATTCGATTTCTTTAAGACTCAAGAAAATTTACAAGACCAAGGTTCAAAACACATTC	2916
Db	9354	ATTGGAATTCACATGTATTAAGTCCGCAAGCTTCAAGACCTGTTGTGATGTGTTT	9413

QY	2917	AAAAAGTTGATACCTTGACTGCTGCTGTTGGTTCAGTTTACAAABAATATTAATCCAA	2976
QY	2917	AAAAAGTTGATACCTTGACTGCTGCTGTTGGTTCAGTTTACAAABAATATTAATCCAA	2976
Db	9414	AAAAAGTGGAGTACCTTGACACCGCTGTGTGGTAGCGGTACAAAGAGTTTATCACCCAA	9473
QY	2977	ACTGCTAACGTTAAGAAAACCGCTGCTTATAGTGTGCTGTAATCTGGTGATTAAGAA	3036
Db	9474	ACAGAGTGATATAAAAAAGCTCTCTAGTGCCTTGAAAGATGCTAATTAACGCTGCTCTTAA	9533
QY	3037	CTGTGCTGCTGCTTAAACCTGCTGCTCCAGCTTAACGATCTGCACCAAAACCTA	3091
Db	9534	GCACGAGTTAAACCAAGCTGCTCCAAACGCTCCAAAGACCAACCAATCCACCACTTA	9588

	00000002				DNA linear	BCT 24-NOV-2000
LOCUS	AE000002	16876 bp				
DEFINITION	Mycoplasma pneumoniae M129 section				15 of 63 of the complete genome.	
ACCESSION	AE000002	U00089				
VERSION	AE000002.2	GI:11379441				
SOURCE						
REMARKS						
ORGANISM	Mycoplasma pneumoniae Mycoplasma pneumoniae Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma. (bases 1 to 16876)					
FEATURES	Himmelreich,R., Hilbert,H., Plagens,H., Pirk,E., Li,B.C. and Hermann,R. Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae Nucleic Acids Res. 24 (22), 4420-4449 (1996)					
TITLE						
JOURNAL						
MEDLINE						
PUBMED	8948633					
REFERENCE	2 (bases 1 to 16876) Dandekar,T., Huynh,N., Regula,J.T., Ueberle,B., Zimmermann,C.U., Andrade,M.A., Boerks,T., Sanchez-Pillado,L., Snell,B., Suyama,M., Yuan,Y.P., Hermann,R. and Bork,P. Re-annotating the Mycoplasma pneumoniae genome sequence: adding value, function and reading frames Nucleic Acids Res. 28 (17), 3278-3288 (2000)					
AUTHORS						
TITLE						
JOURNAL						
MEDLINE	20411492					
PUBMED	10954595					
REFERENCE	3 (bases 1 to 16876) Himmelreich,R., Hilbert,H. and Li,B.-C. Direct Submission Submitted (15-NOV-1996) Zentrum fuer Molekulare Biologie Heidelberg, University Heidelberg, 69120 Heidelberg, Germany					
AUTHORS						
JOURNAL						
TITLE	4 (bases 1 to 16876) Suyama,M., Dandekar,T. and Hermann,R. Direct Submission Submitted (15-JUN-2000) Zentrum fuer Molekulare Biologie Heidelberg, University Heidelberg, 69120 Heidelberg, Germany					
AUTHORS						
JOURNAL						
TITLE	On Nov 27, 2000 this sequence version replaced gi:1673651. This updated annotation replaces the old annotation from reference from 1. The old gene identifiers (WP numbers) according to the original publication by Himmelreich et al. (1996) are given as well as new gene numbering (MPN numbers) based on the origin of replication. Annotation comments and further update data are at http://www.bork.embl-heidelberg.de/Annot/MF/ .					
COMMENT						

```

FEATURES
source
    Location=unclassified
    1..16876
    /organism="Mycoplasma pneumoniae"
    /mol_type="genomic DNA"
    /strain="M129"
    /db_xref="taxon:2104"
    /note="ATCC 29342"
    138..1112
    /gene="orf4"
    /note="synonym: E07-orf324"
    138..1112
    /gene="orf4"
    /note="synonym: E07-orf324"
    138..1112
    /gene="orf4"
    /note="MPN140(new), 015(Himmelfreich et al., 1996)"
    /codon_start=1

```

/transl_table=4
/product="DHH family phosphoesterases"
/protein_id="AAB95662.1"
/db_xref="GI:1673660"
/translation="MNSOVHRKGSIAEVAISIAQHDKIYFHHIRPDGCLGAHQHGLA
RLIOTNFPKROVCGVDPKHNPEMVEVTPKEQITPELMQGLAVIVANVYERIEC
RLDLLDNOFKAVRLIDHNPEDLITNHNFDVASYIAAEOVYVDLVAQKMKLSPAA
TALYLIGITDSNRELISNTSMRTLYGSMILYRAQNIARIKHDELNHTSLKIDFKOYV
FKNEOFONVIVYFVADKKFQKLIKVTPLFCARNYLITANI EQGFIMLPFIEEGNNHRV
ERSNGINVREVALKIGGGGHHQASAVILKSKRDIIRVQDCKQIAV"
1125..6008
/gene="p1"
/note="synonym: E07_orf1627"
1125..6008
/gene="p1"
/note="MPN141(new), 014(Himmelreich et al., 1996)"
/product="ADP1_MYCPN adhesin p1"
/transl_table=4
/codon_start=1
/protein_id="AAB95661.1"
/db_xref="GI:1673659"
/translation="MHQTKTALSSTWILLTATASLTAGTGVGHFTSTTTLRQ
QSYTRDEVALKHNALINRLTPWYRNITSESLPLTGENPGAMALVHDNSKGLTA
GSGSQQTTPDPTTEALASTFTFALRYDLGALRYDLPSKLNRPQTPRDTQGIT
FNFPGFGLSGAAPQOMNEVKNVPEVADSPNPRFALIVPRSVYEEQLORIGL
LPQORTSGONTSTTGAMFGKLVKNAEDTAKNEKLOGEATSGSTSGSGSTORG
GSSGDKVKALKTEVKKSDSDNCOLOLEKNDLANAPIKRSESGSOVQKADDCGT
ALSSSGSGNSNGSPTPWRPLATBOIHKDLPKMSALILITDAPYARKRTAIDVYD
HLDPKAMTANYPSPMTRPKNNHGLMDMKARDVLLQTGFENRRHP EWFDDGQIVAD
NERTGPDVNSENTKQGFQKEADSDKSAIPLFEAYFANIGLWFGQALLVFGNG
HYTKSAHTAPLSIGVPRVNRATGTSATGMPYALLFSGMVKQOTGLKQDFNNR
WEYVPRMAVAGAKFVGRRELYLAGTITMGDTAVPRLVLELSNINLVAOOGGLRE
DLQFTPGGANRPDLIGAMSSSSSSSHNAPYFPHNNDMODRPYONVDAIRKWE
DKNGKDAKIYVRYSGWAMQVYWNKSKLTDPISADPVNNAQPNSPALAILN
ELLALPDVKYKKEKNEFANETERTNOKITVAPTOGTWNSHPSPILSRSTELNIG
SVLDQVLDVYVPIGNGYRYGNHRRGVDITAPQISGSSGISTNTSGSRPLPTSN
IGVGLANVOATLGSGQITWITGSPRRITDQANLIQMTGAGMNDKASGSDENHTK
FTSATMDQOGSGTSGAGNDSLKODNISKSGSLTODGNAIDQOEATNYTLPNL
TTPADMPNALSTFNKNNAOAOFLRGLGSIPIVLNNGSDSNKFOATDOKKSYDL
HSDPTKLNLPAYGEVGLNLPALVPRFGMTRGSGSNTSSPGIGFKIPEDNSK
ATLITPGLAMTPDOVGNLVVSGTTSFQGLGWLVTTFDVKPRAGLYGLQTLGLDSD
ATQRALIMAPRPAARFGSMVNRGLGVEVMDLKVWADOAOSQSTTATARNALP
EHPNALAFQVSYVEASAYKPNNTSSQOTQSNISPLYLHVPKRVQSDKLDLKNIL
DPNOVKTRLROSGTDSHTOPQOSLKTTPVGTSGSNLSVLSGGGAGSGSGGSO
SGVDLSPEVEKSGMLVGLPSTSDGNTSSNAPNTNNGNDVYVGRISSESAAMN
DDVGVIRTPPLALDGEQOTADTPGQVYKFKSPDIDENRLTTHVYTDLFVDYMLY
YDQYIFLEIDIPASVNPKNVRLVVSFDNEOSLGLRLEFKPDQDTQPPNNVQVFN
NGDFLPLTASSQGPQLTSPFNQMPDYLPLAITVPIVIVYLVGLAIGIPMKN
KQALKAGFALSNQKVDLTKAVGVSVEKEIINRTGISQAPKRLKOTSAKGAARPPVP
PKGAPKPPVQPKKPA"
1125..2671
/repeat_family="REPMP4"
/rpt_family="REPMP4"
3171..4971
/rpt_family="REPMP2/3"
6014..9670
/gene="E07_orf1218"
/note="synonym: orf6"
6014..9670
/gene="orf6"
6014..9670
/gene="orf6"
6014..9670
/gene="orf6"
CDS
/note="MPN142(new), 013(Himmelreich et al., 1996)"
/codon_start=1
/transl_table=4
/product="involved in cytaherence"
/protein_id="AAB95660.1"
/db_xref="GI:1673658"
/translation="MKSRLKRLRYLFLPLPTGTLANTYLLLODHNLTLPYPTPT
PLNGLDIVRAALHPSTYELVDMKRVGDKTIALVLSALVRAKEDDTSSDOSNTNP
ALSFPDQESOKALNGSGSSDTSNDSQDFASYVLITFKAAPRATWYFEKIKTALPY
VKQESGSGDQSGNGSKLTKLLODLLEDPVPIYTPNGLARVNGVADQDTVFGSGO
ESSWNSORSDOKLNNPGKAVTGFKLDGRAVRKLNESVPEVPELDTSEGGKDES
SMKNSERTKAENDAPLVGVNGVSGAAGSASSLQNGSNSGLKSLRSAPVSPSPSST

NOTLSLNPAPVGPQAVVSPQAGATAAVSNRTASDTATFSKYLNTFAQLHOMGVIV
PGLKMGNGNGGVAVSRQDATSTNLPHAAAGSQTGILVNGSGSPREPLALPTSORAVV
AGPLRAGNSSEIDALPNVITQTLHTSTADALITNGLVNGSGRPSVSLWYVYGEQ
SGKATMANKTELMDGDKOKOFENOLGRKDDNSKSNLSKNAQGLTPATVLAELN
VVAHDILVEAFAPAGAVGYDMTIDSSASTYNOALAWSTYAGLSDGGYKALVENTALN
GPINGLFTLLDFAVYTPVSGMAGSQNNEDVQTYVPASDQATKAKISLNASPLN
SYGDQSVTVEADLGLNFKLNELERLSPRDEQDLVYGINESLKSARNAQSTSDN
SNKVKMTWNTASHYLVPPVYSANPEACNRRARBOKNGVITSLLESQADGFRANSL
NCTGTGKACVDAPVAPVGRKPNYSANVLYRGVYRANFNPBDTKLLDSTDKNEPISF
STPFRSASAVDLTTLKQVITYAESGLWFFYFDEKFPYDQKQVKNRKYAVIT
VSRTEGIEFEDANTTTLQAPALAVQNGIASSQDDQTLGILPSDFSAVITKQDWT
GKVADIKNTNGNGLEKDDQLSENVKRRNDGLVPIYNGEIVDINGRVDFAVANSYQANL
TKQTVDEVINNPDILOSFFKFPFAPNORAMLVGEKTSPTTLVVKPIEYLDNFGYE
DSKIAGIPLNIDEPSRIIFAGPALPSWVPIVSGSSVGITLILILIGLIGIPMYKVR
KLQDSSFVDVPEFKVPDLITRANGSVYKKIITQTSVIKKAPSALKANNAPKAPVRPA
PVARPPVQPKKPA"
6014..8714
/rpt_family="REPMP5"
9787..10314
/gene="E07_orf175"
9787..10314
/gene="E07_orf175"
/note="MPN143(new), 012(Himmelreich et al., 1996)"
/transl_table=4
/codon_start=1
/product="hypothetical protein"
/protein_id="AAB95659.1"
/db_xref="GI:1673657"
/translation="MSHKDFNGLOABOLSSSPVAKKQSHKRLRALHARYLNHSS
KRLIKHALLHEHNOVLLKEGSPNODMYSKQPGYKSLKLNKLSIGVSKESPK
KRPPYPTKGSKTTPBEAKALQOMQSQKRFHEHNSPLNEVANNPMIQRKQKQA
KRAANRORTYKXRO"
10457..11977
/rpt_family="REPMP2/3"
10888..12129
/gene="E07_orf413"
10888..12129
/gene="E07_orf413"
/note="MPN144(new), 011(Himmelreich et al., 1996)"
/codon_start=1
/transl_table=4
/product="hypothetical protein, see: MPN141"
/protein_id="AAB95658.1"
/db_xref="GI:1673656"
/translation="MGQOGSGTSGAGNPSLQKDTISKSGSLTTODGNATGOEATN
YTNLPNLPITADPNALSTFNKNHRRDLRGLGSIPIVLNNGSDSNKFOATD
QKWSYTDLOSQDKLNLPAVEVGNLNPALVETFGNTRAGSGSNTSSPGIGFKI
PEGNNDKSKATLITPGLAMTPDOVGNLVVSGTSLSPQGLGWLVSFTDFIKPRAGYGLQ
LSGLDSDSDORRELIVAKRMAFRCGSVNRGLRGSVMDLKVWADOAOLAQAATS
EASGSLAPHPNALFOVSYVEASAYSSSTSSGSGSSNTSPYHLIRKRYESTTO
LDQGLKNLDPNNOVTRKLRSGTSDHTQPOSILKTTTPVFGTSSGNIGSVLSGGAGG
GSSGSGSGVDLSPEVRSCH"
12481..13020
/gene="E07_orf179"
12481..13020
/gene="E07_orf179"
/note="MPN145(new), 010(Himmelreich et al., 1996)"
/codon_start=1
/transl_table=4
/product="hypothetical protein, see: MPN013"
/protein_id="AAB95657.1"
/db_xref="GI:1673655"
/translation="MSYSPSLKELITIALQKYTSKNYQSNCKRPPDCKIELSLNGFEE
IKVTPGKTRIVTHKQDQKDKODLMEHLNTPTRKTDLAKSEARINKKELALQI
VWVQGEQIKVHGEOINKLTQVAVEKQKEIEAQOQIQOKVNETINFEVYESIGSIHKRLD
SMGGRDLSMENRDLKLESK"
13338..14135
/gene="E07_orf265"
13338..14135
/gene="E07_orf265"
/gene="E07_orf265"
Query Match 3.24; Score 102.2; DB 1; Length 16876;
Best Local Similarity 55.5%; Pred. No. 1.1e-07;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 06:06:51 ; Search time 798 seconds
(without alignments)
10787.604 Million cell updates/sec

Title: US-09-901-572a-2

Perfect score: 3189
Sequence: 1 atgaatattctcaaaaact.....caaccgctcccaagaataa 3189

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3144.2	98.6	3189	18	DNA encoding antiq
2	3143.4	98.6	10651	22	Vector pCDNA3-Fast
3	124.6	3.9	580073	18	Mycoplasma genital
4	78.4	2.5	4182	16	DNA encoding cyta
5	64.6	2.0	3435	16	Mycoplasma plium a
6	62.6	2.0	4985	24	Anopheles gambiæ
7	60.8	1.9	8771	24	Human immune syste
8	60.2	1.9	3489	21	Kaposi's sarcoma-a

C	9	60.2	1.9	3489	22	AAF82901	Nucleotide sequenc
C	10	60.2	1.9	3489	24	ABA93487	Kaposi's sarcoma-a
C	11	60.2	1.9	32207	20	AAV73805	KSHV LTR DNA (nucl
C	12	60.2	1.9	137507	19	AAV19941	KSHV long unique c
C	13	59	1.9	535	24	ABO42292	Oligonucleotide fo
C	14	59	1.9	535	24	ABO42293	Oligonucleotide fo
C	15	58.4	1.8	640681	24	ABA92787	Buchnera sp. genom
C	16	58	1.8	50000	24	ABA55644	AmPV genome fragm
C	17	55.8	1.7	969	24	ABO39490	Oligonucleotide fo
C	18	55.8	1.7	969	24	ABO39491	Oligonucleotide fo
C	19	54.8	1.7	3837	21	AAV70211	Plasmidium falcipla
C	20	54.6	1.7	14041	22	AAH48024	Internal control B
C	21	53.8	1.7	3582	21	AAV70241	Plasmidium falcipla
C	22	53.4	1.7	1159	21	AAA59240	An EcoRI fragment
C	23	53.2	1.7	987	18	AAV90555	Clostridium perfr
C	24	53.2	1.7	1037	21	AAA59242	Clostridium perfr
C	25	53.2	1.7	1472	21	AAA59241	Exons E, C and A o
C	26	53.2	1.7	8056	25	ABE10246	Exons D, C, B and
C	27	53	1.7	987	18	AAV90554	Haematopoietic cel
C	28	53	1.7	13784	24	ABR40062	Clostridium perfr
C	29	52.4	1.6	3057	20	AAV99555	Human chemically p
C	30	52	1.6	972	20	AAV61754	Nucleic acid seque
C	31	52	1.6	1029	20	AAV61753	B. burgdorferi ant
C	32	52	1.6	5120	22	AAV61752	B. burgdorferi ant
C	33	52	1.6	5120	22	AAV61753	DNA sequence of hu
C	34	51.8	1.6	8056	25	AAV20248	Borrelia burgdorfe
C	35	51.6	1.6	2451	22	AAV07045	Haematopoietic cel
C	36	50.6	1.6	901	23	AB108561	Drosophila melanog
C	37	50.6	1.6	6087	23	AB108560	Drosophila melanog
C	38	50.2	1.6	2010	24	AAV18114	Streptococcus dysg
C	39	50	1.6	2355	24	ABN71061	Streptococcus poly
C	40	50	1.6	2448	24	ABN65054	Streptococcus poly
C	41	50	1.6	50000	24	AB155643	AmPV genome fragm
C	42	49.8	1.6	215561	24	ABN71527	Streptococcus poly
C	43	49.8	1.6	486	22	AAV73507	polyglutamine trac
C	44	49.4	1.5	1887	20	AAV9539	Nucleic acid seque
C	45	49.2	1.5	40324	24	ABO67150	Human angiogenesis

ALIGNMENTS

RESULT 1	AAV75087	AAV75087 standard; DNA; 3189 BP.
ID	AAV75087	
AC	AAV75087	
XX		
DT	27-FEB-1998	(first entry)
XX		
DE	DNA encoding antigenic protein derived from Mycoplasma gallisepticum.	
XX		
KW	Mycoplasma gallisepticum antigen; epitope: monoclonal antibody: bird;	
KW	Mycoplasma infection; viral vaccine; avipoxvirus; herpesvirus; ss.	
XX		
OS	Mycoplasma gallisepticum.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..3189
FT		/tag= a
FT		/product= Mycoplasma_antigen
FT	CDS	1..3189
FT		/tag= b
FT		/codon= (seq: tga, aa: trp)
XX		
PN	WO9724370-A1.	
PD	10-JUL-1997.	
XX		
PF	27-DEC-1996.	96MO-JP03863.
XX		
PR	28-DEC-1995.	95JP-0352754.
XX		

QY 1801 CTAAGTGAAGAGTGTGAGTATTCCTCTAATACCTCATATATAGACACAGGTGAC 1860
 |||||
 Db 1801 CTAAGTGAAGAGTGTGAGTATTCCTCTAATACCTCATATATAGACACAGGTGAC 1860
 QY 1861 ACACGAGAAAGCCGAGAGATCTTCCAAATCTGGCTATTCGATTAATAGTATGATACAT 1920
 |||||
 Db 1861 ACACGAGAAAGCCGAGAGATCTTCCAAATCTGGCTATTCGATTAATAGTATGATACAT 1920
 QY 1921 CAATAGCTTTAGATTTGATGGAATTAATTAATTAATGTTGGGTTAAAGCATCA 1980
 |||||
 Db 1921 CAATAGCTTTAGATTTGATGGAATTAATTAATTAATGTTGGGTTAAAGCATCA 1980
 QY 1981 AGCTTCTTAACCTCAAAATAGACCAATCCAAACGGCTGGAATGTTGCTGCAACACA 2040
 |||||
 Db 1981 AGCTTCTTAACCTCAAAATAGACCAATCCAAACGGCTGGAATGTTGCTGCAACACA 2040
 QY 2041 TACTTAAGATCAAAATAGGATTAAGTACATCTGATTAACCAACCAACCATTC 2100
 |||||
 Db 2041 TACTTAAGATCAAAATAGGATTAAGTACATCTGATTAACCAACCAACCATTC 2100
 QY 2101 GGAACACCTCACCAAGTATTTGATACATCAGCTGATGATCACTTCATCAATTAAGAT 2160
 |||||
 Db 2101 GGAACACCTCACCAAGTATTTGATACATCAGCTGATGATCACTTCATCAATTAAGAT 2160
 QY 2161 ATTAGAACAATCTTCCCTGGTAAACAGTATGATGATCACTTCATCAATTAAGAT 2220
 |||||
 Db 2161 ATTAGAACAATCTTCCCTGGTAAACAGTATGATGATCACTTCATCAATTAAGAT 2220
 QY 2221 AAATCTAGTGTATTAATTAAGATTAAGTACATCAAGTACACCTGATGCTCAAGCTCA 2280
 |||||
 Db 2221 AAATCTAGTGTATTAATTAAGATTAAGTACATCAAGTACACCTGATGCTCAAGCTCA 2280
 QY 2281 TTCAGTCCACAAAGTATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAG 2340
 |||||
 Db 2281 TTCAGTCCACAAAGTATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAG 2340
 QY 2341 AATTCATCTATACAGTAAATCTGCTGATTAAGTACATCAAGTACATCAAGTACATCAAG 2400
 |||||
 Db 2341 AATTCATCTATACAGTAAATCTGCTGATTAAGTACATCAAGTACATCAAGTACATCAAG 2400
 QY 2401 TCTCTGATCATATACCTGCTGATTAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2460
 |||||
 Db 2401 TCTCTGATCATATACCTGCTGATTAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2460
 QY 2461 GAAGGTTCTGCTGATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2520
 |||||
 Db 2461 GAAGGTTCTGCTGATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2520
 QY 2521 GAGTACTTAATCAAAATGGGTTCACTAGTCAAGTGGCTAGAAATCTGTTACAAACCA 2580
 |||||
 Db 2521 GAGTACTTAATCAAAATGGGTTCACTAGTCAAGTGGCTAGAAATCTGTTACAAACCA 2580
 QY 2581 AACTCTTAACAGTATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2640
 |||||
 Db 2581 AACTCTTAACAGTATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2640
 QY 2641 GTGCTTGAATCTGATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2700
 |||||
 Db 2641 GTGCTTGAATCTGATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2700
 QY 2701 TACTTAGATGTAAGTATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2760
 |||||
 Db 2701 TACTTAGATGTAAGTATTAAGTACATCAAGTACATCAAGTACATCAAGTACATCAAGTACAT 2760
 QY 2761 TATTAACAATTTGGGCTTACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 2820
 |||||
 Db 2761 TATTAACAATTTGGGCTTACTTATGATGATGATGATGATGATGATGATGATGATGATGAT 2820
 QY 2821 GGTATTTCTGCAATTAATGATCACTTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTA 2880
 |||||
 Db 2821 GGTATTTCTGCAATTAATGATCACTTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTA 2880

QY 2881 AGAAATTAACAAGACAAAGGTTCAAAACCAATTCACAAAGTATTAAGTATTAAGTATTAAGTATTA 2940
 |||||
 Db 2881 AGAAATTAACAAGACAAAGGTTCAAAACCAATTCACAAAGTATTAAGTATTAAGTATTAAGTATTA 2940
 QY 2941 GCTGTTGTTCAAGTATTAACAAGATTAATTAACCAAACTGCTTAAGTATTAAGTATTAAGTATTA 3000
 |||||
 Db 2941 GCTGTTGTTCAAGTATTAACAAGATTAATTAACCAAACTGCTTAAGTATTAAGTATTAAGTATTA 3000
 QY 3001 GCTGCTTTAGTGTGCTGATTAATCTGCTGATTAAGTATTAAGTATTAAGTATTAAGTATTA 3060
 |||||
 Db 3001 GCTGCTTTAGTGTGCTGATTAATCTGCTGATTAAGTATTAAGTATTAAGTATTAAGTATTA 3060
 QY 3061 GCTCCAGCTTAACCAATCTGCTGCTGATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTA 3120
 |||||
 Db 3061 GCTCCAGCTTAACCAATCTGCTGCTGATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTA 3120
 QY 3121 TCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
 |||||
 Db 3121 TCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
 QY 3181 AAAGATTA 3189
 |||||
 Db 3181 AAAGATTA 3189

RESULT 2
 AAF88025
 ID AAF88025 standard; DNA; 10651 BP.
 XX
 AC AAF88025;
 XX
 DT 16-JUL-2001 (first entry)
 XX
 DE Vector pCDNA3-FasL-IRES-crima DNA.
 XX
 KW Vector; pCDNA3-FasL-IRES-crima; autoimmune disease; apoptosis; crima;
 KW suicide enzyme; antiinflammatory; immunomodulatory; rheumatic;
 KW arthritic; dermatological; gene therapy; rheumatoid arthritis;
 KW systemic lupus erythematosus; Sjogren-Syndrome; polymyositis;
 KW dermatomyositis; polymyalgia; rheumatism; arteritis temporalis;
 KW Crohn's disease; Behcet's disease; colitis ulcerosa; thyroiditis;
 KW autoimmune hepatitis; diabetes mellitus (type I); psoriasis; dermatitis;
 KW pemphigus vulgaris; multiple sclerosis; myasthenia gravis;
 KW viral infection; brain inflammation; transplant rejection; ds.
 XX
 OS Synthetic.
 XX
 PN WO200127254-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 12-OCT-2000; 2000MO-DE03608.
 XX
 PR 12-OCT-1999; 99DE-1048983.
 XX
 PA (SCHW/) SCHWARZMANN F.
 XX
 PI Schwaerzmann F.
 XX
 DR WPI; 2001-282021/29.
 XX
 PT New gene transfer vector useful in the prevention and treatment of
 PT autoimmune and chronic inflammatory disease -
 XX
 PS Claim 9; Page 74-79; 82pp; German.
 XX
 CC This invention describes a novel gene transfer vector (NI) which
 CC comprises (I) a first nucleic acid sequence (I) encoding one or
 CC more ligands that trigger apoptosis; (2) a second nucleic acid sequence
 CC (II) encoding one or more antigens; and optionally (3) a third nucleic
 CC acid sequence (III) encoding one or more anti-apoptosis molecules; and
 CC optionally (4) a fourth nucleic acid sequence (IV) encoding one or more
 CC suicide enzymes. The products of the invention have antiinflammatory,

Db 1806 CTAAGTGAAGAGTGTGAGAAAGTTCTCTAATACTCCATATATATAAGACACAGGTGAC 1865
QY 1861 ACACGAGAAGCCGAGACATCTTCCATCTGCGCTATCTGATATATAGTATGACATT 1920
Db 1866 ACACGAGAAGCCGAGACATCTTCCATCTGCGCTATCTGATATATAGTATGACATT 1925
QY 1921 CAATCAGTTTTAGATTTTGTAGGAATTAGAAATTAATTAAATGTTGGGTTAAAGCATCA 1980
Db 1926 CAATCAGTTTTAGATTTTGTAGGAATTAGAAATTAATTAAATGTTGGGTTAAAGCATCA 1985
QY 1981 AGCTCTTAATCTAAATAGAACCAATCCAAAGGCTGAGAAATGATGCTGCAACANCA 2040
Db 1986 AGCTCTTAATCTAAATAGAACCAATCCAAAGGCTGAGAAATGATGCTGCAACANCA 2045
QY 2041 TACTTAAGATCAACAATTTGATTTAGCTAGAACATCTGATTCACCAACCAAGCATTC 2100
Db 2046 TACTTAAGATCAACAATTTGATTTAGCTAGAACATCTGATTCACCAACCAAGCATTC 2105
QY 2101 GGACACACTCACCAAGTTATTTTCACTATCAGCTGATCACTTCATCAATTAAGANT 2160
Db 2106 GGACACACTCACCAAGTTATTTTCACTATCAGCTGATCACTTCATCAATTAAGANT 2165
QY 2161 ATTAAACAATCTCCGCTGATACGATTAAGTATGCTTCTTATTCACAAATGAAATATAT 2220
Db 2166 ATTAAACAATCTCCGCTGATACGATTAAGTATGCTTCTTATTCACAAATGAAATATAT 2225
QY 2221 AAATCTAGTCTTTATACATTAAGATTAAGTATGATGACTCAAGTACCTGATCAAGCTCA 2280
Db 2226 AAATCTAGTCTTTATACATTAAGATTAAGTATGATGACTCAAGTACCTGATCAAGCTCA 2285
QY 2281 TTCACTCCAAACAAGTTTAAATGACGTTAAATGAATTTGTTGTAATCTTACCTTTATTAGAC 2340
Db 2286 TTCACTCCAAACAAGTTTAAATGACGTTAAATGAATTTGTTGTAATCTTACCTTTATTAGAC 2345
QY 2341 AATTCATCTTACAGTAAGTAAATGCTGCTGATGTTGCTATGCTTTCATCAACAACCTGCT 2400
Db 2346 AATTCATCTTACAGTAAGTAAATGCTGCTGATGTTGCTATGCTTTCATCAACAACCTGCT 2405
QY 2401 TCTCCTGATCATATACTGCTGTAATATACATTTAATACAGAACTTATCTGATATGCTTTT 2460
Db 2406 TCTCCTGATCATATACTGCTGTAATATACATTTAATACAGAACTTATCTGATATGCTTTT 2465
QY 2461 GAAGCTTCTGCTGCTAAGTATACATCTGATTTCTGGGGAACAATCCAAATTCAAACCCGAT 2520
Db 2466 GAAGCTTCTGCTGCTAAGTATACATCTGATTTCTGGGGAACAATCCAAATTCAAACCCGAT 2525
QY 2521 GAGTACTTAATCAAAATGGGTTCACTACTCAAGGCTGAGAAACCTGCTTAAACCAACA 2580
Db 2526 GAGTACTTAATCAAAATGGGTTCACTACTCAAGGCTGAGAAACCTGCTTAAACCAACA 2585
QY 2581 AGCTTCTTAAACAGTTTAACTGATGACTGCTGCTAATGCTGATCACTACGCTGTA 2640
Db 2586 AGCTTCTTAAACAGTTTAACTGATGACTGCTGCTAATGCTGATCACTACGCTGTA 2645
QY 2641 GTGTTGATCTGATGCTAATTTAAACCAACCAACCTCACTTAAAGTTCAAGATCAAA 2700
Db 2646 GTGTTGATCTGATGCTAATTTAAACCAACCAACCTCACTTAAAGTTCAAGATCAAA 2705
QY 2701 TACTTAAGATGCTAATGATGCTAATTTAAAGCAAAATTAATTAAGTATGATCTCT 2760
Db 2706 TACTTAAGATGCTAATGATGCTAATTTAAAGCAAAATTAATTAAGTATGATCTCT 2765
QY 2761 TATTAACAATTTGGGCTTACCTTCATGGGATGAGTACCAAGCAATTTGATGATCAATTA 2820
Db 2766 TATTAACAATTTGGGCTTACCTTCATGGGATGAGTACCAAGCAATTTGATGATCAATTA 2825
QY 2821 GGTATTTCTGCAATTAATGATCACTTAAAGATTAAGTATGCTTAAAGGCTCAAA 2880
Db 2826 GGTATTTCTGCAATTAATGATCACTTAAAGATTAAGTATGCTTAAAGGCTCAAA 2885
QY 2881 AGAAAAATTCAGACCAAGGCTTCAAAACCAACATTTCAAAAAAGTTGATACCTTGACTGCT 2940
Db 2886 AGAAAAATTCAGACCAAGGCTTCAAAACCAACATTTCAAAAAAGTTGATACCTTGACTGCT 2945

QY 2941 GCTGTGGTTCAAGTATTAAGAAAGATTATTACCAAACTGCTAAGCTTAAGAAAAAAGCT 3000
Db 2946 GCTGTGGTTCAAGTATTAAGAAAGATTATTACCAAACTGCTAAGCTTAAGAAAAAAGCT 3005
QY 3001 GCTGCTTAAGTGTGCTGCTAATCTGCTGATTAAGAAACCTGCTGCTCTTAAACCTGCT 3060
Db 3006 GCTGCTTAAGTGTGCTGCTAATCTGCTGATTAAGAAACCTGCTGCTCTTAAACCTGCT 3065
QY 3061 GCTCAGCTTAACCAATCTGCTCAGCAAAAGCTGCTCAGCTTAACCAATCTGCTGCTTAA 3120
Db 3066 GCTCAGCTTAACCAATCTGCTCAGCAAAAGCTGCTCAGCTTAACCAATCTGCTGCTTAA 3125
QY 3121 TCTGTGCGCTTACCAAAACCAACTGCTCCTTAAGCCAGCTGCTCCTCAAAACCAACCGCTCC 3180
Db 3126 TCTGTGCGCTTACCAAAACCAACTGCTCCTTAAGCCAGCTGCTCCTCAAAACCAACCGCTCC 3185
QY 3181 AAGAAATAA 3189
Db 3186 AAGAAATAA 3194
RESULT 3
AAT58840
ID AAT58840 standard; DNA; 580073 BP.
XX
AC AAT58840;
XX
DT 27-MAR-1997 (first entry)
XX
DE Mycoplasma genitalium genome.
M. genitalium; DNA; DNA gyrase; origin of replication;
KM megabase shotgun sequencing method; open reading frame; ORF; ss.
XX
OS Mycoplasma genitalium.
XX
FH key
FT 8552..9184
FT CDS
FT /label= a
FT /note= "Previously identified as MORF-20076, the
FT encoded protein shows 27.59 percentage
FT identity to thymidylate kinase (CDC8)
FT from *Saccharomyces cerevisiae*"
FT 11252..12040
FT /tag= b
FT /label= MG009
FT /note= "Previously identified as MORF-20078, the
FT encoded protein shows 35.43 percentage
FT identity to the *Bacillus subtilis* hypothetical
FT protein covered in accession number
FT GB:D26185_102"
FT 12069..12725
FT CDS
FT /tag= c
FT /label= MG010
FT /note= "Previously identified as MORF-20079, the
FT encoded protein shows 25.73 percentage
FT identity to DNA primase (dnaB) from
FT *Clostridium acetobutylicum*"
FT complement (13570..14247)
FT /tag= d
FT /label= MG012
FT /note= "Previously identified as MORF-20080, the
FT encoded protein shows 31.50 percentage
FT identity to the ribosomal protein S6
FT modification protein (rimK) from *Escherichia*
FT coli"
FT complement (14396..15217)
FT /tag= e
FT /label= MG013
FT /note= "Previously identified as MORF-19823, MORF-20080
FT and MORF-20081, the encoded protein shows 33.04
FT

percentage identity to 5,10-methylene-tetra-
hydrofolate dehydrogenase (fold) from E. coli"

17474..19243
/tag= f
/label= MG015
/note= "Previously identified as MORF-20084, the
encoded protein shows 32.23 percentage
identity to transport ATP-binding protein
(msbA) from E. coli"

26478..27344
/tag= g
/label= MG023
/note= "Previously identified as MORF-20092, the
encoded protein shows 45.96 percentage
identity to fructose-bisphosphate aldolase
(tst) from B. subtilis"

27345..28448
/tag= h
/label= MG024
/note= "Previously identified as MORF-19826 and
MORF-20093, the encoded protein shows 46.84
percentage identity to GTP-binding protein
from E. coli"

36987..38978
/tag= i
/label= MG032
/note= "Previously identified as MORF-20099, the
encoded protein shows 26.82 percentage
identity to ATP-dependent nuclease (adda)
from B. subtilis"

39242..39904
/tag= j
/label= MG033
/note= "Previously identified as MORF-20100, the
encoded protein shows 35.90 percentage
identity to glycerol uptake facilitator
(gltP) from B. subtilis"

complement (39873..40514)
/tag= k
/label= MG034
/note= "Previously identified as MORF-20101, the
encoded protein shows 48.13 percentage
identity to thymidylate kinase (tdk)
from B. subtilis"

40543..41787
/tag= l
/label= MG035
/note= "Previously identified as MORF-20102, the
encoded protein shows 30.71 percentage
identity to histidyl-tRNA synthetase (hlsS)
from Mycobacterium leprae"

complement (44751..46277)
/tag= m
/label= MG038
/note= "Previously identified as MORF-20105, the
encoded protein shows 46.83 percentage
identity to glycerol kinase (gltK)
from E. coli"

complement (46268..47422)
/tag= n
/label= MG039
/note= "Previously identified as MORF-19831 and
MORF-20106, the encoded protein shows 43.20
percentage identity to glycerol-3-phosphate
dehydrogenase (GDR2) from S. cerevisiae"

49377..49643
/tag= o
/label= MG041
/note= "The encoded protein shows 48.86 percentage
identity to phosphotransferase (ptsH) from Mycoplasma
capricolum"

50060..51520

/tag= p
/label= MG042
/note= "Previously identified as MORF-19832 and
MORF-20108, the encoded protein shows 41.92
percentage identity to spermidine/
putrescine transport ATP-binding protein
(potA) from E. coli"

51525..52382
/tag= q
/label= MG043
/note= "Previously identified as MORF-20110, the
encoded protein shows 26.51 percentage
identity to spermidine/putrescine transport
system permease protein (potB) from E. coli"

52366..53220
/tag= r
/label= MG044
/note= "Previously identified as MORF-20111, the
encoded protein shows 29.45 percentage
identity to spermidine/putrescine transport
system permease protein C (potC) from E. coli"

54658..55605
/tag= s
/label= MG046
/note= "Previously identified as MORF-20112, the
encoded protein shows 36.60 percentage
identity to sialoglycoprotease (gcp)
from Pasteurella haemolytica"

complement (56970..58310)
/tag= t
/label= MG048
/note= "Previously identified as MORF-19834,
MORF-20114 and MORF-20115, the encoded protein
shows 43.02 percentage identity to signal
recognition particle protein (fth) from B.
subtilis"

58117..59079
/tag= u
/label= MG049
/note= "Previously identified as MORF-20114 and
MORF-20115, the encoded protein shows 44.78
percentage identity to purine-nucleoside
phosphorylase (deoD) from E. coli"

59083..59754
/tag= v
/label= MG050
/note= "Previously identified as MORF-20117, the
encoded protein shows 83.03 percentage
identity to deoxyribose-phosphate aldolase
(deoC) from Mycoplasma pneumoniae"

complement (64896..65731)
/tag= w
/label= MG056
/note= "Previously identified as MORF-20122, the
encoded protein shows 30.25 percent
identity to the protein disclosed in
GB:D6185.99 from B. subtilis"

complement (65713..66249)
/tag= x
/label= MG057
/note= "Previously identified as MORF-20123, the
encoded protein shows 38.90 percentage
identity to the protein disclosed in
GB:D26185_104 from B. subtilis"

81047..82597
/tag= y
/label= MG067
/note= "Previously identified as MORF-19845, the
encoded protein shows 28.84 percentage
identity to glutamic acid specific protease
(SPase) from Staphylococcus aureus"

91065..91919
/tag= z

CC 20 consecutive residues. Also described: (1) an isolated polynucleotide comprising: (a) a nucleotide sequence encoding the purified Anopheles gambiae olfaction polypeptide; or (b) a nucleotide sequence that hybridises under stringent conditions to a hybridisation probe comprising a 1964, 1239, 1142, 1236, 1194, 1176, 474 or 1206 nucleotide sequence (see ABQ75102 to ABQ75105 and ABQ75110 to ABQ75113) (52), or its complement; and (2) a method for identifying an agent that binds to mosquito olfaction molecules comprising: (a) providing an isolated mosquito olfaction molecule; (b) contacting a test agent with the isolated mosquito olfaction molecule; and (c) detecting specific binding of the test agent to the isolated mosquito olfaction molecule, where the presence of specific binding identifies the test agent as a mosquito olfaction-binding compound. The mosquito olfaction molecules are useful for mosquito management, i.e. controlling this pest and disease vector. A method from the present invention of screening for substances that modulate arrestin-odourant receptor interaction is useful for identifying pest control agents. The present sequence represents Anopheles gambiae Odourant receptor 2 genomic DNA from the present invention.

CC N.B. The features given in figure 4a are tentative and do not directly encode SEQ ID NO:6 (ABP52835).

XX Sequence 4985 BP; 1319 A; 1058 C; 990 G; 1618 T; 0 other;

Query Match

Best Local Similarity 45.5%; Score 62.6; DB 24; Length 4985;

Matches 333; Conservative 0; Mismatches 396; Indels 3; Gaps 3;

OY 253 ATCAATTTAGATAGTTTACTTAACCATTTATGTTTGAATCTAGTATGTTGGT 312
 DB 3070 ATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3011
 OY 313 GGATCAAAAGTAAACAAATAGTTTCA-GATTACACACTAGAGAAATAGATTGATCA 371
 DB 3010 CAACACACACACAAATTAATAGAAATTAATTAATTAATTAATTAATTAATTAATTA 2951
 OY 372 AAGACAAACAGACATATTATGCTGTGTTGTTAATGATGAGCTAACGTTTCAAT-TTAA 430
 DB 2950 TAACATATGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2891
 OY 431 AAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 490
 DB 2890 AAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2831
 OY 491 TTGCTGCTGTATATATCCAGCTCAAGTAAATGATTATCTGATGAGGACTAAATTTA 550
 DB 2830 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2771
 OY 551 ATTTTCAAAACCAAACTCAAGGTGAATGTTAATGACTTATTTAGATGCGCAATCT 610
 DB 2770 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2711
 OY 611 TAC-CCTAAGATTTTACACCAGATGCTTAATTAATTAATTAATTAATTAATTAATTAATTA 669
 DB 2710 AACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2651
 OY 670 AATGACGTCAACACTGCACTGTTGCTTCCCTGGCCAGTGAATGAGTTAGTGAACAATGCT 729
 DB 2650 TATTTATGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2591
 OY 730 GATGATGCAATGTTGATGTTGGGAATGCTCAATTAATTAATTAATTAATTAATTAATTAATTA 789
 DB 2590 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2531
 OY 790 ACTAAACCACTAGCTAGTATTAATCAAAATCTTCAACTTTAATTCAGAGCAATGCGCTGGT 849
 DB 2530 AGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2471
 OY 850 GCAAAACATATAGATTCCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 909
 DB 2470 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2411
 OY 910 TTAGATGAAAAATTTTATTTATTCAGAAATGAGCTGTTGAGAGGATTAATAAATTAATTA 969

DB 2410 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2351
 OY 970 ACATTAGCTACT 981
 DB 2350 AATTTCTGTAAT 2339

RESULT 7

ID ABL33825 standard; DNA; 8771 BP.

AC ABL33825;

DT 26-MAR-2002 (first entry)

DE human immune system associated gene SEQ ID NO: 1798.

XX Human: immune system disease; cytosine methylation; antiasthmatic;

KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;

KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antirheumatic; antidiabetic; antipsoriatic;

KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;

KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

XX 03-JAN-2002.

PF 02-JUL-2001: 2001MO-EP07537.

PR 30-JUN-2000: 2000DE-1032529.

XX 01-SEP-2000: 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation -

PS Claim 1: SEQ ID NO 1798; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

XX diseases. The present sequence is a gene of the invention.

XX Sequence 8771 BP; 2734 A; 59 C; 1660 G; 4318 T; 0 other;

Query Match

Best Local Similarity 1.9%; Score 60.8; DB 24; Length 8771;

Matches 200; Conservative 0; Mismatches 232; Indels 0; Gaps 0;

OY 256 AATTTAGATAGTTTACTTAACCATTTATGCTTATGATCTAATGATGTTGGTGGCA 315
 DB 5053 ATTAATCTATACATACATATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4994
 OY 316 TACAAGTAAACCAATAGTTTCAGATTACACACTAGCAAGAAATGATGTTGATCAAGA 375
 DB 4993 TACTACACAAAAAACCCTTAATAAACATTTATGCAAAATTAATAAATTAACCAATTA 4934
 OY 376 CAAACAAGACATATTATGCTGCTGTTAATGATGAAGCTAACGTTCAATTTAAAAAGA 435

ID	AAA30290/c	standard; DNA; 3489 BP.
XX	AAA30290;	
XX	11-SEP-2000	(first entry)
XX	Kaposi's sarcoma-associated herpesvirus LANA gene.	
XX	Kaposi's sarcoma-associated herpesvirus; KSHV; rhadino virus;	
XX	latency-associated nuclear antigen; LANA; gamma-2 herpes virus;	
XX	Human herpes virus 8; HHV8; rhadino virus cis-acting element; RYCAE;	
XX	Kaposi's sarcoma; primary effusion lymphoma; PEL;	
XX	human immunodeficiency virus; HIV; multicentric Castleman's disease; ds	
XX	Kaposi's sarcoma-associated herpesvirus.	
XX	Key	Location/Qualifiers
XX	CDS	1..3489
XX		/*tag= a
XX		/product= "LANA"
XX	misc_signal	40..50
XX		/*tag= b
XX		/note= "nuclear localisation signal, NLS"
XX	misc_signal	190..210
XX		/*tag= C
XX		/note= "nuclear localisation signal, NLS"
XX	WO200029626-A1.	
XX	25-MAY-2000.	
XX	19-NOV-1999;	99MO-US27508.
XX	19-NOV-1998;	98US-0109422.
XX	21-APR-1999;	99US-0298568.
XX	(KIEF/) KIEF E D.	
XX	(BALL/) BALLESTAS M E.	
XX	(KAYE/) KAYE K M.	
XX	Kieif ED, Ballestas ME, Kaye KM;	
XX	WPI: 2000-387829/33.	
XX	P-PSDB; AAY96255.	
XX	Treating or preventing a disease associated with rhadino virus	
XX	Infection in a mammal which includes Kaposi's Sarcoma and Primary	

PT Effusion Lymphoma -
XX
PS Disclosure; Fig 6; 70pp; English.
XX
XX The present sequence is the Kaposi's sarcoma-associated herpesvirus,
CC (KSHV) latency associated nuclear antigen (LANA) gene. KSHV is also known
CC as Human Herpes Virus 8 (HHV8) and belongs to the rhadno virus, or
CC gamma-2 herpes virus class. The LANA protein is necessary for the
CC efficient persistence of rhadno virus DNA in mammalian cells. Persistent
CC rhadno virus infection is implicated in a variety of diseases e.g. Kaposi's
CC Kaposi's Sarcoma (KS), Primary Effusion Lymphoma (PEL) and multicentric
CC Castleman's disease. In addition, KS is a common malignancy in HIV
CC patients. KSHV persists in host cells in a latent form. One of the few
CC genes expressed from the latent viral DNA is LANA. LANA associates with
CC both human chromosomes and with the rhadno virus cis-acting element
CC (RVCAE), thereby providing a tethering function: the KSHV DNA episome is
CC "tied" to the host chromosomes. This allows the viral DNA to persist in
CC the host cell. The present sequence may be used to screen and identify
CC molecules that inhibit LANA interaction with RVCAE, thereby interfering
CC with the latency cycle of this virus. Potential antiviral treatments for
CC the above mentioned diseases may therefore be based on LANA deregulation.
SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;

Query Match	1.9%	Score 60.2	DB 21:	Length 3489;
Best Local Similarity	61.6%	Fred. No. 0.004;		
Matches 114; Conservative	0;	Mismatches 68;	Indels 3;	Gaps 1;
QY	2998	CCTGCTGGTTTAGGTAGTGTATTAATCTGGTGATAGAAGAAACCTGCTGCTGCTAAACCT	3057	
Db	2236	CATTCCTGTGCGTGCATCCATCCTGCTGCTGCTCAWCCCTGGCTGCGTGCATTCCT	2177	
QY	3058	GCTGCTCCAGCTAAACAATCTGCACCAGCAAGTACGCCAGGCTAACAACACTGGGC--	3115	
Db	2176	GCTGCTGCGTGCATCCGCTGCTGCTCATCTGCTGCTGCTGCTGCTGCATCCGCGTCT	2117	
QY	3116	-CTAAATCTGTTGGCGCTACAAAACCAACTGCTCTGAAGCCAGCTGCTCCAAAACCAAC	3174	
Db	2116	GCTCATTCGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2057	
QY	3175	GCTCC 3179		
Db	2056	GCTGC 2052		
RESULT 9				
AAF82901/c				
ID	AAF82901	standard; DNA: 3489 BP.		
XX	AAF82901:			
DT	29-JUN-2001	(first entry)		
XX				
DE	Nucleotide sequence of KSHV tethering protein, LANA.			
XX				
KW	Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;			
KM	Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;			
KW	KSHV; latency-associated nuclear antigen; LANA; ds.			
XX				
OS	Kaposi's sarcoma associated herpesvirus.			
XX				
FM	Key	Location/Qualifiers		
FT	CDS	1..3489		
FF		/*tag= a		
PN	WO200125484-A2.			
PD	12-APR-2001.			
PF	29-SEP-2000; 2000WO-US26908.			
PR	01-OCT-1999; 99US-0410399.			
XX				

PA (UNMI) UNIV MICHIGAN.
XX
PI Robertson ES, Colter MA;
XX
DR WPI: 2001-281736/29.
XX
DR P-PSDB: AAB62331.
XX
XX A composition for use in gene therapy comprises an expression vector
PT that includes a nucleic acid sequence encoding a nucleic acid binding
PI protein -
XX
XX
XX Disclosure: Fig 9A; 60pp; English.
XX
XX The invention provides a composition comprising nucleic acid, histone H1
CC protein and expression vector operably encoding a protein suitable
CC for tethering the nucleic acid to the histone H1 protein, where the
CC tethering protein is LANA. The composition is useful in aiding the
CC retention of the viral DNA in the host cell. The viral vector encodes a
CC protein suitable for tethering DNA to histone H1. Methods for screening
CC for compounds which are agonistic or antagonistic for the tethering of
CC viral proteins to histone H1 and DNA binding sites are useful for
CC developing the method of viral transfer. The composition has applications
CC to gene therapy, including the treatment of multiple sclerosis,
CC Parkinson's disease, Huntington disease and diabetes. The present
CC sequence represents the nucleotide sequence of the Kaposi's sarcoma
CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
CC antigen (LANA), which acts as a tethering protein.
XX
SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;
Query Match 1.9%; Score 60.2; DB 22; Length 3489;
Best Local Similarity 61.6%; Pred. No. 0.004;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;
OY 2998 CCTGCTGTTTAAAGTGGTGAATCTGTGTAAGAAGTCTGCTGCTTAAACCT 3057
DB 2236 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2177
OY 3058 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3115
DB 2176 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2117
OY 3116 -CTAATCTGGTGGCGCTTACCAAAACCAAGTCTGCTGCTGCTGCTGCTGCT 3174
DB 2116 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2057
OY 3175 GCTCC 3179
DB 2056 GCTGC 2052
RESULT 10
ABA93487/C
ID ABA93487 standard; DNA: 3489 BP.
XX
XX ABA93487;
AC
DT 25-APR-2002 (first entry)
XX
XX Kaposi's sarcoma-associated herpesvirus LANA protein encoding DNA.
DE
XX Kaposi's sarcoma-associated herpesvirus; KSHV; LANA; RVCAE; PEL;
KW KSHV terminal repeat; rhadino virus cis acting element; episome;
KW primary effusion lymphoma; latency-associated nuclear antigen;
KW gene therapy; gene transfer; gene; ds.
XX
XX Human herpesvirus 8.
OS
XX
XX Key 1.3489 Location/Qualifiers
FH CDS
FT /tag= a
FT /product= "LANA protein"
FT /note= "latency-associated nuclear antigen"

XX
PN US6322792-B1.
XX
PD 27-NOV-2001.
XX
XX 21-APR-1999; 99US-0298568.
XX
XX 19-NOV-1998; 98US-109422P.
XX
XX (KIEF/) KIEFF E D.
PA
PI Kieff ED, Ballesstas ME, Kaye KM;
XX
XX WPI: 2002-153769/20.
DR P-PSDB: ABB05621.
XX
XX System for episomal retention of plasmids in mammalian cells, useful in
PT gene therapy, comprises rhadinoviral LANA and RVCAE sequences
PI
PS Claim 1; Fig 6; 27pp; English.
XX
XX The present invention describes a system (A) for maintaining a plasmid
CC as an episome in mammalian cells, comprising the rhadinoviral sequence
CC LANA (latency-associated nuclear antigen) of 3489 base pairs (see
CC ABA93487, S1) expressed in the cell, and the rhadinoviral sequence RVCAE
CC (rhadinoviral cis-acting element) of 801 base pairs (see ABA93488, S2)
CC present in the plasmid. Also describes is a method for maintaining a
CC closed circular DNA in a cell by expressing (S1) in the cells and having
CC (S2) as a cis-acting and maintenance sequence in the DNA. (A) is
CC particularly used in gene therapy (or other gene transfer applications)
CC that uses mammalian cells in which LANA is expressed. (A) improves
CC persistence of gene therapy vectors in cells. The present sequence
CC encodes Kaposi's sarcoma-associated herpesvirus (KSHV, also called human
CC herpesvirus 8) LANA protein, which is used in the exemplification of the
CC present invention.
XX
SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 other;
Query Match 1.9%; Score 60.2; DB 24; Length 3489;
Best Local Similarity 61.6%; Pred. No. 0.004;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;
OY 2998 CCTGCTGTTTAAAGTGGTGAATCTGTGTAAGAAGTCTGCTGCTTAAACCT 3057
DB 2236 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2177
OY 3058 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3115
DB 2176 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2117
OY 3116 -CTAATCTGGTGGCGCTTACCAAAACCAAGTCTGCTGCTGCTGCTGCTGCT 3174
DB 2116 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2057
OY 3175 GCTCC 3179
DB 2056 GCTGC 2052
RESULT 11
AAV73805
ID AAV73805 standard; DNA: 32207 BP.
XX
XX AAV73805;
AC
DT 25-FEB-1999 (first entry)
XX
XX KSHV LUR DNA (nucleotides 105,301-137,507).
DE
XX
XX Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHF; Bcl-2;
KW dihydropyrimidine dehydrogenase; LUR; long unique region; vaccine; prophylaxis;
KW diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;
KW glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2;

KW		interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis
KW		complement-binding protein; glycoprotein; capsid protein IV;
KW		immediate early protein; Kaposi's sarcoma; protective vaccine
KM		Lymphoproliferative disease; leukaemia; splenomegaly; mycosis
KX		HIV immune status; anti-inflammatory agent; therapy; ds.
XX		
OS		Kaposi's sarcoma-associated herpes virus.
FH	Key	Location/Qualifiers
FT	CDS	1142..2794 /*tag= a /product= complement-binding protein
FT	CDS	8699..11236 /*tag= b /product= glycoprotein B
FT	CDS	complement (17261..17875) /*tag= c /product= interleukin 6
FT	CDS	complement (21548..21832) /*tag= d /product= macrophage inflammatory protein II
FT	CDS	complement (27137..27424)
FT	CDS	/*tag= e /product= interferon regulatory factor 1
FT	CDS	28661..29741 /*tag= f /product= protein Tl.1
FT	CDS	complement (58976..60175) /*tag= g /product= glycoprotein M
FT	CDS	complement (69412..69915) /*tag= h /product= glycoprotein L
FT	CDS	complement (88410..88910) /*tag= i /product= interferon regulatory factor 2
FT	CDS	88600..90541 /*tag= j /product= interferon regulatory factor 3
FT	CDS	90173..90643 /*tag= k /product= glycoprotein X
FT	CDS	complement (93636..94127) /*tag= l /product= interferon regulatory factor 4
FT	CDS	complement (111931..112443) /*tag= m /product= capsid protein IV
FT	CDS	complement (123808..127296) /*tag= n /product= immediate early protein
XX		
PN		WO9804576-A1.
PD	05-FEB-1998.	
PP	22-JUL-1997;	97WO-US13346.
XX		
PR	29-NOV-1996;	96US-075669.
PR	25-JUL-1996;	96US-0686243.
PR	25-JUL-1996;	96US-0686349.
PR	25-JUL-1996;	96US-0686350.
PR	25-JUL-1996;	96US-0687253.
PR	25-JUL-1996;	96US-0688814.
PR	05-SEP-1996;	96US-0708678.
PR	10-OCT-1996;	96US-0728323.
PR	13-NOV-1996;	96US-0747887.
PR	13-NOV-1996;	96US-0748640.
XX		
PA	(UYCO) UNIV COLUMBIA NEW YORK.	
XX		
PI	Bohenzky RA,	Chang Y, Edelman IS, Moore PS, Russo JU;
XX		

DR WPI: 1998-130615/12.
XX
XX New nucleic acid encoding Kaposi's sarcoma associated herpes virus
PT proteins - useful for, e.g. detecting levels of HHV8 in, and
PT preparation of vaccines for treatment of, HIV patients
XX
XX Example 2: Page 135-203: 230pp; English.
XX
XX This sequence represents the long unique region and terminal repeat of
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
CC invention which encode KSHV polypeptides selected from: (a) viral
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
CC (e) capsid protein IV encoded by ORF65; and (e) immediate early protein
CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
CC by it, and antibodies (Ab) specific for the proteins are useful for
CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
CC fluids or tissue samples. HHV8 infections can be treated with antisense
CC or triplex forming molecules or agents that bind specifically to the
CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
CC while the protein can be used in protective vaccines. Ab may also be used
CC to differentiate between lymphomas, and HHV8 may be implicated in many
CC other lymphoproliferative diseases such as lymphomas, leukaemia,
CC splenomegaly and mycosis fungoides. Cells and animals containing the
CC nucleic acid are useful for drug screening. HHV8-derived peptides can be
CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
CC can be inhibited with methotrexate. These can also be used to determine
CC the immune status of a patient infected with HIV. HHV8 derived protein
CC viral MIP III may be used as an anti-inflammatory agent for,
CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
CC 81 open reading frames.
XX
XX Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;
SQ
Query Match 1.9%; Score 60.2; DB 19; Length 137507;
Best Local Similarity 61.6%; Pred. No. 0.012;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;
OY 2998 CCTGCTCTTTAGTGGCTGTAATCTGTAATAGCAAGCTGCTGCTGTAAC 3057
DB 125061 CATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125120
OY 3058 GCTGCTCAGCTAAACATCTGACCAAAAGCTAGCTACCAAGCTAAACACTGGGC-- 3115
DB 125121 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125180
OY 3116 -CTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3174
DB 125181 GCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 125240
OY 3175 GCTCC 3179
DB 125241 GCTGC 125245
XX
XX RESULT 13
ABO42292/c
ID ABO42292 standard; DNA; 535 BP.
XX
XX ABO42292:
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 28883.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX

PN WO200218632-A2.
XX
XX PD 07-MAR-2002.
XX
XX PF 01-SEP-2001; 2001WO-EP10074.
XX
XX PR 01-SEP-2000; 2000DE-1043826.
XX
XX PR 05-SEP-2000; 2000DE-1044543.
XX
XX PA (EPIC-) EPIGENOMICS AG.
XX
XX PI Olek A, Piepenbrock C, Berlin K, Guelig D;
XX
XX WPI: 2002-371829/40.
XX
XX PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
XX PS Claim 12: 56pp + sequence listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABO13410-ABO54121 represent genetic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
XX SQ Sequence 535 BP; 98 A; 29 C; 55 G; 353 T; 0 other;
Query Match 1.9%; Score 59; DB 24; Length 535;
Best Local Similarity 48.6%; Pred. No. 0.0041;
Matches 161; Conservative 0; Mismatches 170; Indels 0; Gaps 0;
OY 1203 GAATCAAGTTAAACAGATGATATTTGAAGCTTCAAGCACTGATTAACGTAACAACAA 1262
DB 467 GAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 408
OY 1263 CGGAACACGACACAGCTGATCATCTAGTGTTCACAGAGTGTGGAACAGAAATAC 1322
DB 407 TAAACGAATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGAT 348
OY 1323 TACTACACTTTCGCAACAGTTTCTTAATCTTAAATCTTAAATCTTAAATCTTAAATCT 1382
DB 347 TAAATAATCTAATAATAACGAATTAATAATAGATTAATAATCTAATAATAATAATA 288
OY 1383 TGATAGTAACCAACTCTGCAACAAATAGATGAAGTAATTTGGGAGAGTCTAAGT 1442
DB 287 TACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 228
OY 1443 TATTGAACAGATATATGCTGCAATATACAGATTAGTATTTCAAAATGAATTTCAATAC 1502
DB 227 AACGAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 168
OY 1503 TAATGAGGAAGAACTTATTCGGAACACAAAT 1533
DB 167 TAATTAACGAAACGAAATTAATAATAATAAT 137


```
Db 525047 ATTAACAAAAAATAACACAGAAATTTAAATACATACCATTTCTAAAGTAATTTATT 525106
QY 301 GATGATGTGGTGATACAAAGTAACAAATAGTTTCAGATTACACACTAGCAGAAAT 360
Db 525107 TTAACAAAAAGAGAAAAATATTAATTAACCTATTAATCTCAAAAAAGAAATTAATAAAAA 525166
QY 361 AGATTGATCAAGACAAACAGAGCATATTAATGCTCTGTGGTAAATGATGAAGCTAAC 420
Db 525167 GCAATTGTAAATAGATCTTCAATCTAATAAAAAATTAACAATCAAGAAAGCACA 525226
QY 421 GTTCATTTAAGAAATTAATTAATCAATCAATAGAAATGTAATGTAAGAAACAATTC 480
Db 525227 ATTTATTTCAAAAAAATCAAGATTAATTTTATATTCAGAAAAATTTAAATTAATTT 525286
QY 481 AAGTTGTAATGTGGTGTGATATCCAGCTCAGCTAATTAGATTACTGATGATGGG 540
Db 525287 GTTGAATTAACAACTGATTAACCTTAAATTAACATTCGCAAAATTAAGAGATTTATGATGG 525346
QY 541 ACTAATTTAATTTTACAAACCAAGCTCAAGGTGAATGTTAATGACTTCATTTAGAT 600
Db 525347 TATATAGAAATATTAC--ACAATACTCAACTAAAGAAAAAGAGATATAGTATTTTC 525404
QY 601 GCGCCATCTTACCTAAAGATTTACACCCAGATTGGTATPACTTATACATTCAAGAAAG 660
Db 525405 AAGTCAAAATATAACAACAAGCATATCGATATTATCTAGATTGCATTAATACACCAAG 525464
QY 661 ATCTTACCAATGACGTCACA 682
Db 525465 ATTTTCAAAATATGCTCAAGA 525486
```

Search completed: October 10, 2003, 08:36:31
Job time : 805 secs

THIS PAGE BLANK (USTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 10, 2003, 07:41:45 : Search time 6303 seconds

(without alignments)
12296.839 Million cell updates/sec

Title: US-09-901-572a-2

Perfect score: 3189

Sequence: 1 atgaatatttcctcaaaaact.....caacgcctcccaagaataa 3189

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_hiv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.8	2.2	1101	29	CNS0039G
2	67	2.1	1200	13	BX437758
3	66.4	2.1	1101	29	CNS00EVL
4	65.8	2.1	844	29	BX139987

Result No.	Score	Query Match	Length	ID	Description
5	65.2	2.0	1101	29	CNS0039G
6	64.6	2.0	829	29	BX173672
7	64.2	2.0	773	29	CNS01WVC
8	63.8	2.0	1200	13	BX447454
9	61.8	1.9	470	29	FR0018463
10	61.8	1.9	480	29	BZ643398
11	61.8	1.9	843	29	BZ643413
12	61.8	1.9	1225	29	CNS0161D
13	61.6	1.9	735	29	CNS04NSM
14	60.6	1.9	641	28	AQ946120
15	60.2	1.9	1101	29	CNS00EVE
16	60	1.9	941	13	BUI26589
17	60	1.9	1056	13	BX415058
18	59.8	1.9	450	29	FR0025683
19	59.8	1.9	488	12	BJ396861
20	59.8	1.9	575	29	BZ780846
21	59.8	1.9	615	28	BH766948
22	59.8	1.9	994	13	BX414650
23	59.2	1.9	619	29	FR0006944
24	59.2	1.9	1201	13	BX458623
25	59	1.9	317	12	BJ328366
26	58.8	1.8	619	29	FR0013713
27	58.8	1.8	759	29	CNS060XV
28	58.8	1.8	813	29	BX143161
29	58.8	1.8	836	28	BH491109
30	58.8	1.8	1007	29	CNS06X9S
31	58.4	1.8	1126	13	BX446391
32	58.2	1.8	332	14	C92788
33	58.2	1.8	641	12	BM181884
34	58	1.8	257	29	FR0026883
35	57.8	1.8	1101	29	CNS0100X
36	57.4	1.8	1101	29	CNS00ESI
37	57.2	1.8	427	29	BZ422321
38	57.2	1.8	649	12	BM160056
39	57.2	1.8	1101	29	CNS001FB
40	56.8	1.8	639	29	CNS038CX
41	56.8	1.8	989	29	CNS028H4
42	56.6	1.8	1101	29	CNS00EVL
43	56.4	1.8	596	29	FR0019514
44	56.4	1.8	1201	13	BX461105
45	56.2	1.8	501	29	FR0048173

ALIGNMENTS

RESULT 1
CNS0039G/c
LOCUS
DEFINITION
BACR08K10 of RCT-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.
ACCESSION
AL063921.1 GI:4941778
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RCT-98 library from *Drosophila melanogaster* (fruit fly), genomic survey sequence.
AL063921
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; *Drosophilidae*; *Drosophila*.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley *Drosophila* Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of

FEATURES

•

...

1

1
;

1

1

1

2

DISC

VERS I

SOURCE

AUTHORS

COMMENT

FEATURES

ings

1
2
3
4
5

•

10

1

4

5

Db 801 WTTTWTAAAAAATTTMMAGARAAKATWTWMAAAAAAAMFRAAAATATATTTT 860
 QY 612 ACCTA-AAGATTACACCCAGATGTGATTAATCACTTCAAGAAAGACTTCCCA 670
 Db 861 ATATATTAARAAAAATTAATTAARAAARRRKGAATAAAAAATTAATTTWTATTA 920
 QY 671 ATGAGCTCAACACTGCAGTTGTCCTTGCCAGTAGAGTACGTAGACAAATGCTG 730
 Db 921 AAATTTTAMAMWMAAAAMAAAMAAAGAAARAGAMAAATTTTATTAATATRWMA 980
 QY 731 ATGATGGAGTGTGATTTGCGAATGTCAATAATAACAGATCTATGCTCAAA 790
 Db 981 AAAAATTTTATTTTWTARAAAMWMDTTTATTTATTTATWTATAGAAATATADATWRA 1040
 QY 791 CT-----AAACCACTACTGATATCAAAATCCTTCAACTTTAATTCAGGCAATGC 844
 Db 1041 TTATNADAAARAAAGAAATKADAGATAAATTTTATTTTAAAAAAGAAAAA 1100
 QY 845 CTGTGCAACAACTAGATACGATTCATGTAATGTAAGCATAGATTAATAACATCTT 904
 Db 1101 TTTTAAAAAARAAAAAATRAAAAAAAGAAAAAAGAAKADAAAMAGAAATTA 1160
 QY 905 TCCATTTAGTGAATAA 921
 Db 1161 AAAAAAAGMRAARAW 1177

RESULT 3
 CNS05EVL 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
 DEFINITION BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069706 GI:4949849
 VERSION AL069706.1
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain y²; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
 1..1101
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR29B23"
 /clone.lib="RPCI-98"
 /note="end : T7"
 BASE COUNT 419 a 91 c 60 g 299 t 232 others
 ORIGIN

Query Match 2.1%; Score 66.4; DB 29; Length 1101;
 Best local Similarity 35.1%; Pred. No. 0.018;
 Matches 224; Conservative 104; Mismatches 300; Indels 10; Gaps 3;

QY 4 AATATTTCTAAAAAAGCTTAAGATTAATACATGATAGTGATGATGCTGATTTGGAGCT 63
 Db 471 MAATWTWMAAAAMAAATTAATTAATWMAAAAMAAAMAAATTTTMMWMTATTTWTWMAW 530
 QY 64 CTGTGCTTCGCAAGCTTTGGCTTTAAGCAATCGAATAGAGTAACGTAACAGCATTA 123
 Db 531 TATATWMAAAAAAAMAAATAATTTAAAMWATAATTAATTAATTAATTAATW 590
 QY 124 GTTAATCAAGCAAGCAAGCGATGCTAATTCCTGTAAGCTGACGCTTGCAAAAT 183
 Db 591 TATATATWTAATTAATTAATTAATTAATTTTATTAATTAATTAATTAATTAAT 650
 QY 184 GGTTCGCTGTCAATACAGCTTCTAGAGATGATGATTAATTAACAGCAAGCTAAT 243
 Db 651 TATTAATATATWTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 710
 QY 244 GCAACAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 303
 Db 711 AATAATATATWTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 770
 QY 304 GATTGCTGCGATCAACAAGTAACAATAAGTTGCTGATTAACACTAGCAAGTAAGA 363
 Db 771 TATWMAATWMAAMWMTATATWMAATTAATTAATTAATTAATTAATTAATTAAT 828
 QY 364 TTTGATCAAAAGACAAACAAAGCAATTAATGCTGCTGTTGTTATGATGAAGCTA 423
 Db 829 TATWMAATWTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 888
 QY 424 CATTTAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 483
 Db 889 TTTTWTWMAAMWMTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 943
 QY 484 TTTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
 Db 944 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1000
 QY 544 AATTTTAATTTTCAAAACCAACTCAAGCTGAATGTTATTAATTAATTAATTAAT 603
 Db 1001 AATATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1060
 QY 604 CCAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 641
 Db 1061 TTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1098

RESULT 4
 BX139987 844 bp DNA linear GSS 13-MAR-2003
 LOCUS Dario rerio genomic clone DKEY-99E7, genomic survey sequence.
 DEFINITION BX139987
 ACCESSION BX139987
 VERSION BX139987.1 GI:27971314
 KEYWORDS GSS.
 SOURCE Dario rerio (zebrafish)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 844)
 AUTHORS Humphray, S.J., Huckle, E. and Durham, J.L.
 JOURNAL Direct Submission
 TITLE Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humphray@sanger.ac.uk unpublished
 COMMENT This sequence was generated from the T7 end of BAC 99E7. 99E7 is
 part of the Dariokey BAC library created by R. Plasterk and N.V.
 Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.

FEATURES
 source
 1..844
 Location/Qualifiers

```

/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEX-99E7"
/tissue_type="testis"
/notes="vector pindigBAC-536"

```

```

BASE COUNT      406 a      49 c      81 g      308 t
ORIGIN
Query Match      2.1%; Score 65.8; DB 29; Length 844;
Best Local Similarity 45.2%; Pred. No. 0.022;
Matches 241; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

```

```

47 TACCTGTTTGGAGCTCTGTTGTCGACGTTGGCTTTAGCAATCAGATAGATA 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 TTCCAGCTCATTTATGCTTTTATTTACTTTGATCATTAATTAATTAATTA 371
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 ACGATACACGCAATAGTAAATCAAGCAAGACGCTAGCTAATCTGTAGACTG 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 CAGGCTTGGCAAAATGCTGCTTCAATACAGTCTTACAGATGTTGATTA 226
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
227 TTATACACGCAATAGTAAATCAAGCAATTAATTAATTAATTAATTA 286
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
492 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 551
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
287 GTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 611
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
552 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 671
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
347 CACTACACGCAATAGTAAATCAAGCAATTAATTAATTAATTAATTA 731
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
612 ATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
407 ATGATGACGCAATAGTAAATCAAGCAATTAATTAATTAATTAATTA 844
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
672 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA
467 GAAACAAATTTCAATTTGTAATGCTGCTGTAATTAATTAATTAATTA 526
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
732 ACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
527 TTACTGATGAGGCAATTAATTAATTAATTAATTAATTAATTAATTA 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
792 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA

```

```

RESULT 5      1101 bp      DNA      linear      GSS 03-JUN-1999
CNS0039G
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence JET3 end of BAC #
BACR08K10 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL063921
VERSION
AL063921.1 GI:4941778
KEYWORDS
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

```

```

REFERENCE
AUTHORS
TITLE
Direct Submision
JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila

```

melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Kammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Source

```

1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_id="RPI-98"
/notes="end : JET3"
Location/Qualifiers

```

```

BASE COUNT      201 a      64 c      131 g      202 t      503 others
ORIGIN

```

```

Query Match      2.0%; Score 65.2; DB 29; Length 1101;
Best Local Similarity 19.1%; Pred. No. 0.03;
Matches 129; Conservative 264; Mismatches 282; Indels 1; Gaps 1;

```

```

324 AAACAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 383
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
419 AAACAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 478
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
384 AGCATATTAATGCTGTTGATGATGAGGCTTACGTTGATTAATTAAT 443
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
479 AAATATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTT 538
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
444 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 503
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
539 TAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
504 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 563
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
598 TTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
564 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 623
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
658 TAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
624 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 683
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
718 TTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
684 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 743
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
778 TAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 837
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
744 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 803
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
838 TTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 897
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
804 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 863
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
898 TTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 957
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
864 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 923
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
958 TTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1017
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 TAATCAATAGTTCACATTAACACACTAGCAATAGATTGATCAACAAACAG 983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1018 TTAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1077
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
984 TAATTTGCAACAGCAAC 999
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db	1078	MAWHHHHXYAHYHHAH	1093
RESULT 6			
LOCUS	BX173672	829 bp	DNA linear GSS 13-MAR-2003
LOCUS	BX173672/c		
DEFINITION	Danio rerio genomic clone DKer-150M6, genomic survey sequence.		
ACCESSION	BX173672		
VERSION	BX173672.1	GI:28005377	
KEYWORDS	GSS.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Danio rerio		
REFERENCE	Eukariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
AUTHORS	1 (bases 1 to 829)		
TITLE	Humphray, S.J., Hucklee, E. and Durham, J.L.		
JOURNAL	Direct Submission		
COMMENT	Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphraysanger.ac.uk Unpublished		
FEATURES	This sequence was generated from the SP6 end of BAC 150M6. 150M6 is part of the Daniokey BAC Library created by R. Plastek and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_rerio/.		
source	Location/Qualifiers		
	1..829		
	/organism="Danio rerio"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:7955"		
	/clone="DKer-150M6"		
	/issue_type="Testis"		
	/note="vector pindigobAC-536"		
BASE COUNT	260 a 103 c 32 g 434 t		
ORIGIN			
Query Match	2.0%: Score 64.6; DB 29; Length 829;		
Best Local Similarity	45.5%: Pred. No. 0.038;		
Matches 229; Conservative	0; Mismatches 274; Indels 0; Gaps 0;		
OY	85	TTTAAAGCAATGAGTTAAGAGTAAAGAGTAAACGCGAATGTTAATCAAGCAAGAACGTA	144
DB	669	TTGAAACATTGGAAGAAAAACGTAATATATATTAATGAATATTAATTAATGATGAT	610
OY	145	GATGCTAATTCCTGTAGACTTCGACAGCTTGGACAAATGGCTGCTTCAATACAGTT	204
DB	609	GATGATGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	550
OY	205	CTTGAAGATGTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	264
DB	549	AATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	490
OY	265	AGTTTCTAATACCATTAATATATGATGATGATGATGATGATGATGATGATGATGATGAT	324
DB	489	AGAAATATATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT	430
OY	325	AAACAAATAGTTTCAGATTACACACTGACAGAAATGATTTGATCAAGCAAGCAAGA	384
DB	429	AATTAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	370
OY	385	GCATATTAATGCTGCTGTTGGTTAATGATGAGAGCTAAGCTTCAATTAATTAATTAAT	444
DB	369	GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	310
OY	445	AATCAATTAATTAATTTGTAATTAAGAAACAACTTCTAAGTTTGTAAATGGTGGTGT	504
DB	309	AATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	250
OY	505	AATCAAGCTCAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	564
DB	249	GATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	190
OY	565	ACTCAAGCTGAATTTGTTAATGA	587

[illegible]

QY	1420	ACTAATTTGGGCAGATCTCTTAACGTTATTGGAACGAGAATATATGCTGAATCCAGGTTGGT	1479
Db	345	ACTACTCTACTACTACTACTGCAMACACGATACATCTACTCTACTACTATGCTCTACAT	404
QY	1480	ATTCAAAATGAAATTCATTAATCAATATGACAGGAAGTAATTTATCCGAACACAAATTTGGGT	1539
Db	405	ACTGCACACTACGCAACTGTCACACTCTCAACGCMCTACTACCAACACAACTACACAC	464
QY	1540	GTTCGGTTTACTTCAACAGGTTCAAGAGTAAGTTTAAAGAGCTTCTTATTAAGGGTATCA	1599
		:::	:::
Db	465	ACMCACACTACTGCTACACAACTACTACTCMACTGCTACTACTACTACTGCTACTTCT	524
QY	1600	CGTCCAACT	1608
Db	525	ACTACTACT	533

RESULT 8	1200 bp	mRNA	linear	EST 22-MAY-2007
LOCUS	BX447454	Homo sapiens	CDNA clone	CSOCP003Y119
DEFINITION	BX447454	5-PRIME, mRNA sequence.		
ACCESSION	BX447454			
VERSION	BX447454.1	GI:31018746		
KEYWORDS	EST.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens (human)			

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 1200)
Li, W. B., Gruber, C., Jesssee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope

BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6207.r For more information about this cluster, see <http://www.genoscope.cns.fr/cgibin/cluster.cgi?seq=CSOAA22D03QPi&cluster=6207.r>. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSOAA22D03QPi.

FEATURES	Location/Qualifiers
Source	1. 1200

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0CAP003YH19"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="Vector: PCWVSPT_6; 1st strand cDNA was primed
with a NotI-Oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the PCWVSPT 6 vector."

```

BASE COUNT	465 a	222 c	150 g	363 t
ORIGIN				
Query Match	2.0%;	Score 63.8;	DB 13;	Length 1200;
Best Local Similarity	47.6%;	Pred. No. 0.057;		
Matches 188;	Conservative	0;	Mismatches 207;	Indels 0;
				Gaps 0;

[illegible]

QY	343	TACACACTAGCAGAAATAGATTGGATTCACAAAGACAAACCAAGGACATATTGCTGTG	402
Db	926	GAAAAAATATAAAAATTAATTTGAAAAAATAATTAATTTTATTTATGAAATTAA	985
QY	403	GTTAATGATGAGAGCTTAAGCTTATAAAGATTAATTACTAAGTAAATGAAATGCT	462
Db	986	TTTAAAAATATAAAAATTTAAAAAATTTTGAACAATAAATTTGGAA	1045
QY	463	AATAGAACACAAATTCCTAAGTTGGTAATGGTGGTGTGATATCCAGCTACGTAAT	522
Db	1046	ATTAAATATGAGATTTAAATTTGGATTAATTTGAAAAATTTACATTTGATTAATTA	1105
QY	523	AGATTACTAGATGATGGAGCACTAATTTAATTTTACAACCACTCAAGGTGAATTGT	582
Db	1106	GAAATATATGATTAATTTGAAAAAATAATGAATTAATAAAAAAATAAATTAATTTTT	1165
QY	583	AATGACTTCATTTAGATCGGCATCTACTACTA	617
Db	1166	AAATTTAAATTTAATACATTTATTTTTAAAAA	1200

RESULT 9				
FE0018463/c				
LOCUS	FE0018463	470 bp	DNA	linear
DEFINITION	F. rubripes GSS sequence, clone 016E10ac6, genomic survey sequence.			
ACCESSION	AF011359			
VERSION	AF011359.1	GI:2676793		
KEYWORDS	GSS: genome survey sequence.			

ORGANISM Takifugu rubripes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Takifugu.
1 (bases 1 to 470)

AUTHORS: Williams, G. and Brenner, S.
 TITLE: Direct Submission
 JOURNAL: Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
 COMMENT: Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@pump.mrc.ac.uk
 Vector: Bluescript II KS
 V-type: phagemid
 PRIMER: KS
 DESCRI: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.

FEATURES	Location/Qualifiers
source	1. .470

Query Match	1.9%	Score 61.8	DB 29	Length 470
Similarity	51.5%	Pred. No. 0.12		
BASE COUNT	149 a	79 c	150 g	65 t
ORIGIN				27 others
	/organism="Takifugu rubripes" /mol_type="genomic DNA" /db_xref="taxon:31033" /clone="016E10ac6" /clone_1lb="cosmid 016E10"			

[illegible][illegible]

OY	3100	GCTTAACCACTGGGCGCTTAATCTGGGCGGCTCAACAAACCAACTGCTCCTAACCAAGCT	3159
Db	165	GCTACTGCTGCTGCTACTACTGCTGCTACTGCTACTACTGCTACTGCTGCTGCTACTACT	106
OY	3160	GCTCCAAACCAACCAACCGCTCCCA	3181
Db	105	ACTGCTACTACTGCTGCTGCTA	84
RESULT 10			
B2643398		480 bp	DNA
LOCUS			linear
DEFINITION	OGAOW577C.ZM.0.7.1.5_KB zea mays genomic clone ZMBMa011J17,		GSS 29-JAN-2003
ACCESSION	B2643398		
VERSION	B2643398.1		GI:28104877
KEYWORDS	GSS.		
SOURCE	zea mays		
ORGANISM	zea mays		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD 1 (bases 1 to 480)		
AUTHORS	Whitelaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T., Resnick ,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek ,R.W., Nunberg,A., Robbins,D. and Lakey,N. Consortium for Maize Genomics		
TITLE	Unpublished		
JOURNAL	Other_GSSs: OGAOW577M		
COMMENT	Contact: Cathy Whitelaw		
	TIGR		
	9712 Medical Center Drive, Rockville, MD 20850, USA		
	Tel: 301-838-5843		
	Fax: 301-838-0208		
	Email: whitelaw@tigr.org		
	Seq primer: TF		
	Class: sheared ends.		
FEATURES			
source	Location/Qualifiers		
	1..480		
	/organism="Zea mays"		
	/mol_type="genomic DNA"		
	/strain="B73"		
	/db_xref="taxon:4577"		
	/clone="ZMBMa011J17"		
	/clone_id="ZM.0.7.1.5_KB"		
	/note="Vector: pBCKS-; Site 1: HincII; 0.7-1.5 kb		
	methylation filtered genomic DNA library"		
BASE COUNT	96 a 160 c 65 g 159 t		
ORIGIN			
Query Match	1.9%;	Score 61.8;	DB 29; Length 480;
Best Local Similarity	52.5%;	Pred. No. 0.12;	
Matches 135;	Conservative 0;	Mismatches 122;	Indels 0;
	Gaps 0;		
OY	2923	GTTGATACCTTGACTGCTGCTGTTGGTTCAGTTACAGAAGATTATTACCAACTGCT	2982
Db	74	GCTACTACTACTACTGCTGCTGCTACTACTACTGCTACTGCTACTACTACTACTGCTGCT	133
OY	2983	AAGCTTAAGAAAAAACCCTGCGCTTTAGTCTGTTAATCTGGTGATGAAGAAACCTGCT	3042
Db	134	ACTACTACTACTACTGCTGCTGCTACTACTGCTACTGCTACTGCTACTGCTACTACTACT	193
OY	3043	GCTGCTACTTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	3102
Db	194	GCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACT	253
OY	3103	AAACCAACTGGGCGCTAAATCTGTTGGCGCTTACAAACCAACCACTGCTCCTAACCCAGCTCT	3162
Db	254	GCTACTACTGCTACTACTACTACTACTACTACTACTGCTACTACTACTACTACTACTACT	313
OY	3163	CCAAACCAACCGCTCC	3179
Db	314	ACTGCTACTACTGCTAC	330

[illegible]

RESULT 14	
AQ946120/c	
LOCUS	AQ946120 641 bp DNA linear GSS 27-JAN-2000
DEFINITION	Sheared DNA-46J23.TR Sheared DNA Trypanosoma brucei genomic clone
ACCESSION	Sheared DNA-46J23, genomic survey sequence.
VERSION	AQ946120
KEYWORDS	AQ946120.1 GI:6769385
SOURCE	GSS.
ORGANISM	Trypanosoma brucei Trypanosoma brucei Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma. 1 (bases 1 to 641) El-Sayed,N., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C., Gerrard,C., leech,Y., de Jong,P., Ullv.E., Melville,S., Donelson,J., Fraser,C. and Adams,M. Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library Unpublished Other_GSSs: Sheared DNA-46J23.TF Contact: Najib M. El-Sayed Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0200 Fax: 301 838 0208 Email: nelsayed@tigr.org Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through ATCC. Sheared DNA end sequences search page: http://www.tigr.org/tdb/mdb/tbdb/ . Seq primer: M13-Reverse Class: Shotgun.

[illegible][illegible]

LOCUS	1101 bp	DNA	linear	GSS 03-JUN-1999
RESULT 15				
CNS00FVE/c				
DEFINITION	CNS00FVE melanogaster genome survey sequence TET3 end of BAC:			
ACCESSION	BACR32P18 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
VERSION	AI071298			
KEYWORDS	AI071298			
SOURCE	AI071298.1	GI:4951138		
ORGANISM	GSS.			
REFERENCE	Drosophila melanogaster (fruit fly)			
AUTHORS	Drosophila melanogaster			
TITLE	Eukaryote: Metazoa: Arthropoda: Hexapoda: Insecta: Pterygota:			
JOURNAL	Neoptera: Endopterygota; Diptera: Brachycera; Muscomorpha;			
	Ephydroidea: Drosophilidae; Drosophila.			
	1 (bases 1 to 1101)			
	Genoscope.			
	Direct Submission			
	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage			
	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr			
	- Web : www.genoscope.cns.fr)			
	Determination of this BAC-end sequence was carried out as part of a			
	collaboration with the Berkeley Drosophila Genome Project (BDGP).			
	The BDGP is constructing a physical map of the Drosophila			
	melanogaster genome using these BACs. For further information			
	please see http://www.fruitfly.org The BDGP Drosophila			
	melanogaster BAC library was prepared by Kazutoyo Oseguwa and			
	Aaron Mammoser in Pieter de Jong's laboratory in the Department of			
	Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,			
	NY. The library is named RPC1-98 and was constructed by partial			
	EcoRI digestion of Drosophila DNA provided by the BDGP from the			
	isogenic strain Y2: cn bw sp, the same strain used for the BDGP's			
	PI and EST libraries. A more detailed description of the library			
	and how to order individual BAC clones, the entire library, or			
	filters for hybridization from the BACPAC Resource Center can be			
	found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .			
FEATURES	Location/Qualifiers			
SOURCE	1..1101			
	/organism="Drosophila melanogaster"			
	/mol_type="genomic DNA"			
	/db_xref="taxon:7227"			
	/clone="BACR32P18"			
	/clone_lib="RPC1-98"			
	/note="end : TET3"			
BASE COUNT	267 a	223 c	31 g	184 t 396 others
ORIGIN				
Query Match	1.9%	Score 60.2:	DB 29:	Length 1101:
Best Local Similarity	29.8%	Prod. No. 0.23:	245:	Indels 8: Gaps 2:
Matches 165:	Conservative 135:	Mismatches		

Query Match	3.9%	Score 12.6	DB 4	Length 580073
Best Local Similarity	63.5%	Pred. No. 1,9e+18		
Matches 190	Conservative 0	Mismatches 109	Indels 0	Gaps 0
QY	2770	TTTTGGCGCTTACCTTCATTCATGGGTAGTGGCTTACAGCAATTCGTAGTACATTTAGTATTCCTT	2829	
Db	228703	TTTGGCAGCACTCCCTCCGACGGGTGATCCCTGTATCAGTAGGCTTCTTCATGGTGGGATCTTG	228765	
QY	2830	GCAATATATCATCTCTTAGGATTTACTATTCGGTATTCCTTTAAAGCTCAAGAAGAAATTA	2889	
Db	228763	TTTATCTTTTAGTCTCTTAGGACTTGGGATTTGGGATCCCAATGTACAGGTTAAAGAAATC	228822	
QY	2890	CAAGCAAAAGGTTCAAAAACAACATTTCAAAAAAGTTGATACCTTGACTGCTGCTTGGT	2949	
Db	228823	CAAGATGCACTCTTGTGTTAATGTCTTTAAAAAGGTTGATTAACATCCACAACTGCTGGGT	228823	
QY	2950	TCAGTTTACAAGAAGATTTTATCCCAACTGCTAACGTTTAGAAGAAAAAACCTGCGCTTTA	3009	
Db	228883	AGTGTGTACAAAAAAGATTTTATACCCAACTGGTGTGGTGAAGAAAAAGCACCTTAGTGCAATTG	228944	
QY	3010	GGTGTGCTGATAAATCTGTGTGATTAAGAACCCTGCTGCTGTCTTAAACCTGCTCTCAGC	3068	
Db	228943	AAAGCTGTAAATCCTAGTGTATAAAAAACCTGCTGCTTTTAAAAACCACTGTTCACAC	229001	
RESULT	2			

```

US-07-973-257-1
: Sequence 1, Application US/07973257
: Patent No. 5378820
: GENERAL INFORMATION:
: APPLICANT: Keeler, Jr. Calvin L.
: APPLICANT: Dohms, John E.
: TITLE OF INVENTION: Gene Encoding Cytohesin
: TITLE OF INVENTION: Protein of Mycoplasma Gallisepticum and Its Use
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Connolly and Hutz
: STREET: 1220 Market Street
: CITY: Wilmington
: STATE: Delaware
: COUNTRY: U.S.A.
: ZIP: 19899
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM/PC or Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/973,257
: FILING DATE: 19921109
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA: No. 5378820e
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4182 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double stranded
: TOPOLOGY: linear
: MOLECULE TYPE: Genomic DNA
: HYPOTHEetical: No
: ANTI-SENSE: No
: ORIGINAL SOURCE:
: ORGANISM: Mycoplasma Gallisepticum
: STRAIN: S6
US-07-973-257-1

Query Match 2.5%; Score 78.4; DB 1; Length 4182;
Best Local Similarity 98.8%; Pred. No. 1,2e-08;
Matches 79; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGCAATATTTCTTAAAAAACTTAAAGTTATTCATTGATAGTGGATTAGCTGTATTGGA 60
Db 4103 ATGCAATATTTCTTAAAAAACTTAAAGTTATTCATTGATAGTGGATTAGCTGTATTGGA 4162
|||||
OY 61 GCTCTGGTCTGCAAGCTT 80
Db 4163 ACTCTGTGTTCTGCAAGCTT 4182
|||||

RESULT 3
US-08-728-323A-1/c
: Sequence 1, Application US/08728323A
: Patent No. 5948676
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Bohenzky, Roy A.
: APPLICANT: Russo, James J.
: APPLICANT: Edelman, Isidore S.
: APPLICANT: Moore, Patrick S.
: TITLE OF INVENTION: Immediate Early Protein From Kaposi's
: TITLE OF INVENTION: Sarcoma Associated Herpesvirus, DNA
: TITLE OF INVENTION: Encoding Same And Uses Thereof
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham LLP
: STREET: 1185 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.

```

```

      ZIP: 10036
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/728,323A
      FILING DATE:
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
      NAME: White, John P.
      REGISTRATION NUMBER: 28,678
      REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSC/SKS
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 212-278-0400
      TELEFAX: 212-391-0525
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3489 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..3489
      US-08-728-323A-1

Query Match          1.9%, Score 60.2; DB 2; Length 3489;
Best Local Similarity 61.6%; Pred. No. 0.00017;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

QY    2998 CCTGCTGCTTAGTGGCGTGGAATPCTGGTGATAAGAAACCTGCTGCTGTAAACT 3057
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    2236 CATCCCTGCTGCTGCCTGCATCCTGCTGCTGCTGCTCATCTGCTGCTGCTCATCT 2177

QY    3058 GCTGCTCAGCTAAACATCTGCACCAAAAGCTAGCTACCAGCTAAACAACCTGGGC-- 3115
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    2176 GCTGCTGCTGCTCATTCCTGCTGCTGCTGCTATCCTGCTGCTGCTGCTCATCT 2117

QY    3116 -CTAATCTGGTGGCGCTACAAAACCAACTGCTCTTAAGCCAGCTGCTCCAAAACCAAC 3174
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB    2116 GCTATCCTGCTGCTGCCGCTCATCTGCTGCTCATCTGCTGCTGCTGCTGCTGCT 2057

QY    3175 GCTCC 3179
      || | |
DB    2056 GCTGC 2052

RESULT 4
US-09-298-568-1/c
Sequence 1, Application US/09298568
Patent No. 6322792
GENERAL INFORMATION:
APPLICANT: Kieff, Elliott D.
APPLICANT: Ballestar, Mary E.
APPLICANT: Kaye, Kenneth M.
TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
FILE REFERENCE: 16412-1000IR
CURRENT APPLICATION NUMBER: US/09/298,568
CURRENT FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
EARLIER FILING DATE: 1998-11-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1
```

[illegible]

```

RESULT 5
US-09-410-399-1/c
: Sequence 1, Application US/09410399
: Patent No. 6482587
: GENERAL INFORMATION:
: APPLICANT: Robertson, Erle S.
: APPLICANT: Collier, Murray A.
: TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
: TITLE OF INVENTION: to Genomic Host DNA
: FILE REFERENCE: UM-03778
: CURRENT APPLICATION NUMBER: US/09/410,399
: CURRENT FILING DATE: 1999-10-01
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3489
: TYPE: DNA
: ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-1

```

RESULT 6
 US-08-770-379-20
 : Sequence 20, Application US/08770379
 : Patent No. 5849564
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Chang, Yuan
 :
 : APPLICANT: Bohenzky, Roy A.
 :
 : APPLICANT: Russo, James J.
 :
 : APPLICANT: Edelman, Isidore S.
 :
 : APPLICANT: Moore, Patrick S.
 :
 : TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED

1 TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
2
3 NUMBER OF SEQUENCES: 20
4
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: Cooper & Dunham LLP
7 STREET: 1185 Avenue of the Americas
8 CITY: New York
9 STATE: New York
10 COUNTRY: U.S.A.
11
12 ZIP: 10036
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.30
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/770,379
22
23 FILING DATE:
24
25 CLASSIFICATION: 435
26
27 ATTORNEY/AGENT INFORMATION:
28 NAME: White, John P.
29 REGISTRATION NUMBER: 28,678
30 REFERENCE/DOCKET NUMBER: 52342
31
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (212) 278-0400
34 TELEFAX: (212) 391-0525
35
36 INFORMATION FOR SEQ ID NO: 2:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 32207 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: double
41
42 TOPOLOGY: linear
43
44 MOLECULE TYPE: DNA (genomic)
45
46 US-08-770-379-20

[illegible]

RESULT 7
 US-08-757-669A--20
 : Sequence 20, Application US/08757669A
 : Patent No. 6183751
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Chang, Yuan
 : APPLICANT: Bohenzky, Roy A.
 : APPLICANT: Russo, James J.
 : APPLICANT: Edelman, Isidore S.
 : APPLICANT: Moore, Patrick S.
 : TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
 : TITLE OF SEQUENCES: SEQUENCES AND USES THEREOF
 : NUMBER OF SEQUENCES: 20
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Cooper & Dunham LLP
 : STREET: 1185 Avenue of the Americas
 : CITY: New York
 : STATE: New York

```

COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,669A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-F
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 32207 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-757-669A-20

Query Match      1.9%; Score 60.2; DB 3; Length 32207;
Best Local Similarity 61.6%; Pred. No. 0.00039;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

QY 2998 CCTGCTGCTTTAGTGGCTGGTAAATCTGGTATAGAAACCTGCTGCTGCTAAACCT 3057
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19761 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19820
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3058 GCTGCTCCAGCTAAACCATCTGCACAAAAGCTAGCTACGCTAAACCAACTGGGC-- 3115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19821 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3116 -CTAATCTGGTGGCGCTACAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19881 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3175 GCTCC 3179
    |||||
DB 19941 GCTGC 19945

RESULT 8
US-09-230-371A-20
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIOQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; TITLE OF INVENTION: USUS THEREOF
; FILE REFERENCE: 45185-C-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230,371A
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20

```

```

Query Match      1.9%; Score 60.2; DB 4; Length 32207;
Best Local Similarity 61.6%; Pred. No. 0.00039;
Matches 114; Conservative 0; Mismatches 68; Indels 3; Gaps 1;

QY 2998 CCTGCTGCTTTAGTGGCTGGTAAATCTGGTATAGAAACCTGCTGCTGCTAAACCT 3057
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19761 CATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19820
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3058 GCTGCTCCAGCTAAACCATCTGCACAAAAGCTAGCTACGCTAAACCAACTGGGC-- 3115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19821 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19880
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3116 -CTAATCTGGTGGCGCTACAAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19881 GCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3175 GCTCC 3179
    |||||
DB 19941 GCTGC 19945

RESULT 9
US-09-142-584-5
; Sequence 5, Application US/09142584
; Patent No. 6403094
; GENERAL INFORMATION:
; APPLICANT: Tildall, Richard W.
; APPLICANT: Williamson, Ethel D.
; APPLICANT: Havard, Helen L.
; APPLICANT: Oyston, Petra C.F.
; APPLICANT: Payne, Dean W.
; TITLE OF INVENTION: CLOSTRIDIUM PERFRINGENS VACCINES
; FILE REFERENCE: 124-665
; CURRENT APPLICATION NUMBER: US/09/142,584
; PRIOR FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/GB97/00660
; EARLIER FILING DATE: 1997-03-11
; EARLIER APPLICATION NUMBER: GB 9605222.0
; EARLIER FILING DATE: 1996-03-12
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (136)..(987)
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: (1)..(32)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(984)
; OTHER INFORMATION: "n" at positions 451-453 represent a, t, c, g or
; OTHER INFORMATION: other
US-09-142-584-5

Query Match      1.7%; Score 53.2; DB 4; Length 987;
Best Local Similarity 44.3%; Pred. No. 0.0044;
Matches 208; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY 363 ATTGATCAAGACAAACAGACATATTATGCTGTGTTGTTATGATGAAGCTAACCT 422
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 ATCTAATGAAATGTCACAAAAGCTTCTTATGATTAATGATACATTAAATGGAAGG 176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 423 TCATTAAAGAATTAATTAATCAATTAATGAATGATGAACCAATTTCTAA 482
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 177 AAGATATATACAAATATTAATTAATTAAGAGAAATGAAATATTTCTAATGCTAT 236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 483 GTTTGTAATGCTGCTGTTGATTAATCAAGCTACGTAATTAATTAATTAATGATGGAG 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


Db 1099 CTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1158
QY 3050 CTAACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3109
Db 1159 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1218
QY 3110 CTGGGCTTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3169
Db 1219 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1278
QY 3170 CAACGCTGC 3179
Db 1279 CTGCTGCTGC 1288

RESULT 13
US-09-142-584-3
; Sequence 3, Application US/09142584
; Patent No. 6403094
; GENERAL INFORMATION:
; APPLICANT: Tilbail, Richard W.
; APPLICANT: Williamson, Ethel D.
; APPLICANT: Hayward, Helen L.
; APPLICANT: Oyston, Petra C.F.
; APPLICANT: Payne, Dean W.
; TITLE OF INVENTION: CLOSTRIDIUM PERFRINGENS VACCINES
; FILE REFERENCE: 124-665
; CURRENT APPLICATION NUMBER: US/09/142,584
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/GB97/00660
; EARLIER FILING DATE: 1997-03-11
; EARLIER APPLICATION NUMBER: GB 9605222.0
; EARLIER FILING DATE: 1996-03-12
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Clostridium perfringens
; FEATURE:
; NAME/KEY: mat-peptide
; LOCATION: (136)..(987)
; FEATURE:
; NAME/KEY: misc.signal
; LOCATION: (1)..(32)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(984)
; US-09-142-584-3

Query Match 1.7%; Score 53; DB 4; Length 987;
Best Local Similarity 44.6%; Pred. No. 0.0048;
Matches 209; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

QY 363 ATTGTGCAAGCAAGCAAGCAAGCAATATATGCTGCTGCTGCTGCTGCTGCTGCTG 422
Db 117 ATCTAATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 176
QY 423 TCATTAAAAAGATTAATCTACTACTCACTCAATTAATTAATTAATTAATTAATTA 482
Db 177 AAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 236
QY 483 GTTGTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542
Db 237 GGCATATTTGATAGTTACTACTTAATTAATTAATTAATTAATTAATTAATTAATTA 296
QY 543 TAAATTAATTTTAAACAACAACAAGCTCAAGTGAATGTTAAAGACTTCAATTTAGATGC 602
Db 297 TAAAGTTGAATTAATGAGAACCATCAATGAATTAATCTGAAGATGTTATGTTGAAA 356
QY 603 GCCAATCTTAACCTAAAGATTACACCAAGATTGTTAACTTAATTAATTAATTAATTA 662
Db 357 AGCTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 416

QY 663 CTTACCAATGACGCTCAACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
Db 417 TAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 476
QY 723 AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 782
Db 477 AACTGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 536
QY 783 TGTCTAAACTAAACACTACTGATTAATCAAAATCTTCAACTTTTAAAT 831
Db 537 TTTTGCAAAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 585

RESULT 14
US-09-601-198-55
; Sequence 55, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Helmer, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
; LENGTH: 3057
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
; US-09-601-198-55

Query Match 1.6%; Score 52.4; DB 4; Length 3057;
Best Local Similarity 45.6%; Pred. No. 0.01;
Matches 220; Conservative 0; Mismatches 261; Indels 1; Gaps 1;

QY 754 AATGCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 813
Db 1175 AATCTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1234
QY 814 AATCTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 873
Db 1235 TATCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1294
QY 874 TGAATGTCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 933
Db 1295 CTTACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1354
QY 934 GATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 993
Db 1355 AATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1414
QY 994 AGCAACGAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1053
Db 1415 GCATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1474
QY 1054 GATTCAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1113
Db 1475 ATTCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1534
QY 1114 GGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1172
Db 1535 GGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1594
QY 1173 CCAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1232

Db 1595 CTTATTCTTTTATGTAATATCCATCTCTTCATCAAGATTACTAATATCATGTGTATA 1654
QY 1233 TT 1234
Db 1655 TT 1656

RESULT 15

US-09-142-584-1
; Sequence 1, Application US/09142584

; Patent No. 6403094

; GENERAL INFORMATION:

; APPLICANT: Tilbail, Richard W.

; APPLICANT: Williamson, Ethel D.

; APPLICANT: Hayward, Helen L.

; APPLICANT: Oyston, Petra C.F.

; APPLICANT: Payne, Dean W.

; TITLE OF INVENTION: CLOSTRIDIUM PERFRINGENS VACCINES

; FILE REFERENCE: 124-665

; CURRENT APPLICATION NUMBER: US/09/142,584

; EARLIER FILING DATE: 1998-09-11

; EARLIER APPLICATION NUMBER: PCT/GB97/00660

; EARLIER FILING DATE: 1997-03-11

; EARLIER APPLICATION NUMBER: GB 9605222.0

; EARLIER FILING DATE: 1996-03-12

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: Patentln Ver. 2.0

; SEQ ID NO 1

; LENGTH: 987

; TYPE: DNA

; ORGANISM: Clostridium perfringens

; FEATURE:

; NAME/KEY: mat_peptide

; LOCATION: (136)..(456)

; FEATURE:

; NAME/KEY: misc_signal

; LOCATION: (1)..(32)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(984).

; US-09-142-584-1

Query Match

1.68; Score 51.4; DB 4; Length 987;

Best Local Similarity 44.38; Pred. No. 0.011;

Matches 208; Conservative 0; Mismatches 261; Indels 0; Gaps 0;

QY 363 ATTTGATCAAGACAAAGACATATTATGCTGCTGTTAATGATGAAGCTAAGT 422
Db 117 ATCTAATGAATGTCCAAAAGCTTCTTATGATTAATGATGATTAATGATGAAG 176
QY 423 TCATTAAAAAGATTATCTAAGTCAATAGATGTAATAGAAACAATCTTAA 482
Db 177 AAGATATTAATCAAAATATTAATTAAGAGATGAAATATTAATCTTAATGCTAT 236
QY 483 GTTTGATTTGCTGCTGATATATCCAGCTCAAGTATTAATTAATGATGATGAG 542
Db 237 GGCATATTTGATTAAGTCTAATTAATCCAAAGAAATGATTTTATTAATATCC 296
QY 543 TAAATTTAATTTACAAACCAACTCAAGTGAATTTAATGACTTCATTTTATGC 602
Db 297 TAAAGTGAATTAAGTGAAGACCAATCAATTAATCTGAAGATGTTATATGTTGAAA 356
QY 603 GCCAATCTTACCTTAAGATTACACCCAGATGTTAATTAATTAATCAATCAAGAT 662
Db 357 AGCTCTCTTAAGTAAATGATCAACAAAGAAATTAATTAATCAATCAATCACTTG 416
QY 663 CTTACCAATAGACGACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
Db 417 TAAATTAATCTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 476
QY 723 AAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 782

Db 477 AACTGCTAAGTTACTGTTCTTTTAAATGAAGAGATATCAATTAATTAATAG 536
QY 783 TGCTCAAACTAAACACCTACTGATATCAAAATCCCTTCACTTTTAA 831
Db 537 TTTTGAATTAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 585

Search completed: October 10, 2003, 11:49:06
Job time : 183 secs

THIS PAGE BLANK (USPTO)